
بسم الله الرحمن الرحيم

دانشکده مهندسی برق
پایان‌نامه کارشناسی ارشد مهندسی برق مخابرات

بازسازی تصاویر درک شده از روی داده‌های تصویربرداری رزونانس مغناطیسی عملکردی مبتنی‌بر شبکه‌های عصبی

نگارنده : سعیده پورقاسمیان

استاد راهنما :
دکتر حسین خسروی

دی ۱۳۹۹

سپاس گزاری...

پروردگارا مرا یاری کن تا دانش اندکم نه نردبانی باشد برای فزونی و تکبر و غرور، نه حلقه ای برای اسارت و نه دستمایه ای برای تجارت، بلکه گامی باشد برای تجلیل از تو و تعالی ساختن زندگی خود و دیگران.

قبل از هر چیز، خداوند بزرگ را به خاطر لطفی که همواره شامل حال من نموده شاکرم. سپس، از زحمات استاد محترم راهنما، جناب آقای دکتر حسین خسروی که نه تنها به عنوان استاد بلکه همچون همکاری در تمام مراحل انجام این تحقیق از رهنمودها و نگاه های بی دریغ ایشان بهره مند شده ام، تشکر و قدردانی می کنم. همچنین از پدر و مادر و خواهر عزیز، دلسوز و مهربانم که برای من آرامش روحی و آسایش فکری فراهم نمودند تا با حمایت های همه جانبه در محیطی مطلوب، مراتب تحصیلی و نیز پایان نامه درسی خود را به نحو احسن به اتمام برسانم، سپاسگزاری می نمایم. همچنین زحمات جناب آقای مهندس عارف زینی وندنژاد را در راستای اختصاص ساعت های طولانی، بحث و تبادل نظر در مورد موضوع تحقیق بنده، که همواره برای من الهام بخش ایده و دیدگاهی تازه نسبت به موضوع بوده است، سپاس می گویم. در نهایت نیز، تلاش و لطف جناب آقای مهندس حمید باباشاهی و سرکار خانم مهندس مریم فریور را در جهت فراهم ساختن امکانات لازم برای پیش برد این مطالعه را ارج می نهم.

تعهد نامه

اینجانب **سعیده پورقاسمیان** دانشجوی مقطع کارشناسی ارشد رشته **مهندسی برق_مخابرات سیستم** دانشکده **مهندسی برق** دانشگاه صنعتی شاهرود نویسنده پایان‌نامه **بازسازی تصاویر درک شده از روی داده‌های تصویربرداری رزونانس مغناطیسی عملکردی مبتنی‌بر شبکه‌های عصبی** تحت راهنمایی **دکتر حسین خسروی** متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است .
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافت‌های آن‌ها) استفاده شده است ضوابط و اصول اخلاقی رعایت شده است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده است.

تاریخ ۹/ ۱۰ /۱۳۹۹

سعیده پورقاسمیان

چکیده:

در چند دهه اخیر، تحقیقات در زمینه رمزگذاری و رمزگشایی فعالیت‌های مغز از روی داده‌های تصویربرداری رزونانس مغناطیسی عملکردی (fMRI) دستاوردهای چشم‌گیری داشته است. با این وجود، رمزگشایی فعالیت‌های مغز به‌منظور بازسازی تصاویر طبیعی درک شده توسط انسان، همچنان یک مسئله قابل بررسی است. چرا که، این تصاویر حاوی اطلاعات متنوع و پیچیده‌ای هستند.

در این پژوهش، رویکردی بر مبنای شبکه‌های خودرمزگذار رقابتی (AAEs)، به منظور رمزگشایی فعالیت‌های مغز ارائه می‌شود. از این‌رو، ابتدا یک شبکه عصبی AAE، با بیش‌از ۱۱ هزار تصویر طبیعیِ رویت نشده توسط سوژه‌ها، آموزش داده می‌شود. فضای‌نهان این ساختار، توصیفی معنادار از هر تصویر را ارائه می‌دهد. سپس نگاشتی از داده‌های fMRI به فضای‌نهان شکل می‌دهیم، که بر اساس آن داده‌های fMRI به کدهای نهان هم بُعد با کدهای نهان شبکه AAE، تبدیل می‌شوند. به هنگام تستِ مدل رمزگشایی، از نگاشت آموزش دیده، استفاده می‌کنیم و الگوهای fMRI هر سه سوژه انسانی را به کدهای نهان تبدیل می‌کنیم، در نهایت کدها با استفاده از ساختار رمزگشای شبکه AAE، به بازنمایی‌های قابل درکی از تصاویر تست، بدل می‌شوند. مدل رمزگشایی پیشنهادی، با استفاده از داده‌های مجزا، آموزش دیده و تست می‌شود.

به‌منظور سنجش کمی عملکرد روش پیشنهادی، دو معیار شباهت ساختاری چندمقیاسه (SSIM) و ضریب همبستگی (CC) به‌ازای تصاویر بازسازی شده، گزارش می‌شوند. نتایج به‌دست آمده حاکی از عملکرد موفق و امیدبخش رویکرد پیشنهادی است. به‌طوری که در دو سوژه آزمایشی میانگین معیار ضریب همبستگی روی تصاویر تست، بیش از ۰/۷ گزارش شده است.

واژه‌های کلیدی: رمزگذاری و رمزگشایی فعالیت‌های مغز، تصویربرداری رزونانس مغناطیسی عملکردی، شبکه خودرمزگذار رقابتی، شباهت ساختاری چندمقیاسه، ضریب همبستگی.

فهرست مطالب

فهرست شکل‌ها

فهرست علائم اختصاری

LGN	Lateral Geniculate Nucleus
AI	Artificial Intelligence
DNN	Deep Neural Networks
VAEs	Variational Auto-Encoders
GANs	Generative Adversarial Networks
AAEs	Adversarial Auto-Encoders
fMRI	functional Magnetic Resonance Imaging
MRI	Magnetic Resonance Imaging
CAT	Computed Axial Tomography
PET	Positron Emission Ttomography
EEG	Electroencephalography
MEG	Magnetoencephalography
SQUID	Superconducting Quantum Interface Device
HRF	Hemodynamic Response Function
BOLD	Blood Oxygenation level Dependent
CNR	Contrast to Noise Ratio
SNR	Signal to Noise Ratio
CNNs	Convolutional Neural Networks
AEs	Auto-Encoders
DBNs	Deep Belief Networks
RBM	Restricted Boltzmann Machine
SAE	Stacked Auto-Encoder
DAE	Denoising Auto-Encoder
SAE	Sparse Auto-Encoder
CAE	Convolutional Auto-Encoder
SGD	Stochastic Gradient Descent
MVPA	Multi-Voxel Pattern Analysis
BCCA	Bayesian Canonical Correlation Analysis
PGM	Probabilistic Graphical Model
DGMM	Deep Generative Multi-view Model
EPI	Echo Planar Imaging
SS	Single Shot
TR	Repetition Time
TE	Echo Time
Ms-SSIM	Multi-scale Structural Similarity
SSIM	Structural Similarity
CC	Correlation Coefficient
GD	Gradient Descent

PCA	Principal Component Analysis

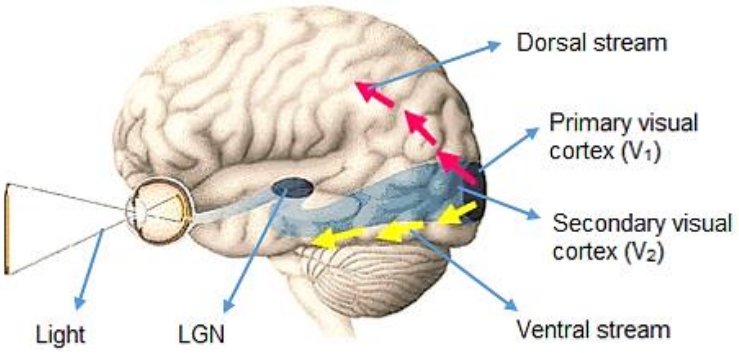
۱- فصل اول

مقدمه

۱-۱- مقدمه

سامانه بینایی انسان از یک سری نواحی غشایی چندگانه که به صورت سلسله مراتبی قرار گرفته‌اند، تشکیل شده است [۱].

درک بینایی انسان، یعنی فرآیندی که به واسطه آن از محیط اطراف شناخت پیدا می‌کنیم، به محض برخورد نور با شبکیه چشم آغاز می‌گردد. این گیرنده‌ها نور را به سیگنال‌های الکتروشیمیایی تبدیل می‌کنند. سیگنال‌های تولید شده از مسیر عصب بینایی، پس از انتقال به نیم کره مقابل، به ناحیه هسته خمیده جانبی [۱](LGN) در تالاموس [۲] (نهنج) می‌رسند. در آن نقطه پس از پردازش‌های اولیه، به سمت قشر اولیه بینایی، همان ناحیه V_1 [۳] حرکت می‌کنند. سپس اطلاعات بینایی به ناحیه V_2 [۴] رفته و در آن ناحیه به دو مسیر مجزا تقسیم می‌شوند: مسیرهای شکمی [۵] و پشتی [۶] (شکل ۱-۱).



شکل ۱-۱: مسیرهای انتقال اطلاعات بینایی از شبکیه چشم به قشر اولیه و ثانویه بینایی و مسیرهای پشتی و شکمی.

در پژوهش‌های پیشین نشان داده شده است که عمده اطلاعات مربوط به ماهیت اشیا در نواحی شکمی پردازش می‌شوند، و ناحیه پشتی مسئول پردازش اطلاعات رنگ و موقعیت اشیا است.

---

[۱] Lateral Geniculate Nucleus

[۲] Thalamus

[۳] Primary visual cortex (V_1)

[۴] Secondary visual cortex (V_2)

[۵] Ventral stream

[۶] Dorsal stream

از این رو نواحی شکمی و پشتی به ترتیب به عنوان نواحی"چیستی[۱]" و "کجایی[۲]" نیز شناخته می‌شوند [۲]. البته استقلال کامل دو مسیر بیان شده، در پژوهش‌های اخیر به چالش کشیده شده است [۳, ۴]. به گونه‌ای که، ارتباطات پیش‌رو[۳]، بازگشتی[۴] و بازخوردی[۵] متعددی در بین نواحی مختلف مغز درنظر گرفته شده است. درواقع، انسان‌ها از طریق یک نگاه، تصویری غنی از محیط اطراف خود به دست می‌آورند. و به سرعت آن را درک می‌کنند [۵]. آن‌ها قادرند بسیاری از اشیاء پیرامون خود را علی‌رغم داشتن حالت‌ها و اندازه‌های گوناگون و همچنین زوایای دید مختلف، بدون کوچک‌ترین مشکلی، بازشناسی و تفکیک کنند [۲]. اطلاعات بینایی به صورت یک بازنمایی بصری در مغز انسان شکل می‌گیرند. آن‌گاه، در نتیجه‌ی بر هم‌کنش‌های پیچیده، ادراک و تمام فعالیت‌های شناختی بعدی انسان که مرتبط با بینایی است، بر روی این بازنمایی‌هایِ شکل گرفته، صورت می‌پذیرد. قرن‌هاست که محققان در تلاش‌اند عملکردهایی از قشر مغز، که امکان درک و کشف محیط بصری را برای انسان فراهم می‌سازند، درک و رمزگشایی کنند. از رایج‌ترین روش‌های پرداختن به این موضوع، بهره‌گیری از روش‌های یادگیری عمیق در چنین تحقیقاتی است.

این مغز است که انسان را قادر می‌سازد این چنین به سادگی دنیای بصری اطراف خود را بشناسد. اگر چه این امر برای انسان بدون صرف زمان زیادی صورت می‌پذیرد، اما یک فرآیند محاسباتی پیچیده و بسیار مشکل است. رمزگشایی و شناخت چگونگی رخداد این فرآیند توسط انسان، پاسخگوی پرسش‌ها و خلاءهای بسیاری در حوزه‌های علوم اعصاب شناختی[۶] و محاسباتی[۷] است، که در نهایت می‌تواند در جهت توسعه ماشین‌های هوشمند شبیه به انسان، کارآمد باشد.

---

[۱] What pathway

[۲] Where pathway

[۳] Feedforward connections

[۴] Feedback connections

[۵] Recurrent connections

[۶] Cognitive neuroscience

[۷] Computational neuroscience

۱-۲- بیان مسئله

از جمله حوزه‌های مرتبط با علوم اعصاب محاسباتی می‌توان به هوش‌مصنوعی (AI)[۱] اشاره کرد؛ به‌طوری که پیشرفت‌های انفرادی اخیر در هر دو زمینه‌ی هوش‌مصنوعی و علوم اعصاب محاسباتی، به راهبردهای جدیدی جهت رشد مشترک هر دو حوزه منجر شده است [۸–۶]. به این شکل که؛ از طرفی بر اساس پیشرفت‌های علمی در زمینه شناخت سامانه بینایی انسان، مدل‌های محاسباتی مصنوعیِ الهام گرفته شده از مغز، در حوزه هوش‌مصنوعی معرفی شده‌اند. مدل‌های مطرح شده؛ به عنوان مثال شبکه‌های عصبی عمیق (DNN)[۲] ؛ توانسته‌اند به عملکرد قابل توجهی در درک تصاویر و فیلم‌ها، دست‌یابند [۱۱–۹]. از سوی دیگر، تطابق و مقایسه چنین مدل‌هایی با سامانه بینایی انسان، به درک و شناخت عمیقی از این‌که مغز چگونه اطلاعات بینایی را بازنمایی می‌کند منجر شده است [۱۵–۱۲]. به این ترتیب می‌توان گفت علوم اعصاب محاسباتی موجب رشد هوش-مصنوعی شده و برعکس، هوش مصنوعی نیز موجب رشد علوم شناختی می شود.

درک سیستم بینایی انسان نیازمند مدل‌های محاسباتی است. این مدل‌ها شامل فرضیات پایه در زمینه محاسبه و یادگیری عصبی هستند [۷]. مدل‌هایی که واقعا منعکس کننده کار مغز در زمینه بینایی طبیعی باشند، باید بتوانند فعالیت مغز را با توجه به هرگونه ورودی بصری توضیح دهند (رمزگذاری مغز [۳]) [۱۶] و فعالیت مغز را به منظور پی‌بردن به ورودی بصری، رمزگشایی کنند (رمزگشایی مغز[۴]) [۱۶]. یک رویکرد پژوهشی امیدوارکننده در زمینه رمزگذاری و رمزگشایی مغز، شامل ادغام روش‌های یادگیری عمیق، در این دسته از تحقیقات است. مدل‌های مولد عمیق، مانند شبکه‌های خودرمزگذار متغیر (VAEs)[۵] [۱۷, ۱۸] و شبکه‌های مولد رقابتی (GANs)[۶] [۱۹] در

---

[۱] Artificial Intelligence

[۲] Deep Neural Networks

[۳] Brain encoding

[۴] Brain decoding

[۵] Variational Auto-Encoders

[۶] Generative Adversarial Networks

زمینه تولید تصاویر به موفقیت‌های بزرگی رسیده‌اند. اخیرا تحقیقات در زمینه بازسازی تصاویر بصری، با استفاده از مدل‌های عمیق تولیدی، گسترش یافته است [۲۵–۲۰]. در پژوهش پیش‌رِو نیز، رویکردی بر مبنای شبکه‌های خودرمزگذار رقابتی (AAEs)[۱]، به منظور رمزگشایی مغز ارائه می‌شود. بدین منظور از داده‌های تصویربرداری رزونانس مغناطیسی عملکردی (fMRI)[۲] به عنوان بازنمایی-هایی از فعالیت‌های مغز استفاده می‌شود. چرا که تحولات اخیر نشان داده است، اندازه‌گیری فعالیت-های مغز به صورت داده‌های fMRI، به تفسیر مطالب ذهنی انسان، از جمله آنچه فرد مشاهده می‌کند، به یاد می‌آورد، تصور می‌کند و حتی در رویا می‌بیند، کمک می‌کند.

در زمینه رمزگذاری و رمزگشایی مغز، مطالعات علوم اعصابِ متداول، از الگوهای مصنوعی یا تصاویر استاتیک برای شناسایی یا بازسازیِ بازنمایی‌های عصبی اجسام یا دسته‌های بصری مجزا استفاده می‌کنند [۱۶, ۲۶, ۲۷]. این در حالی است که، تصاویر طبیعی حاوی اطلاعات متنوع‌تر و پیچیده‌تری هستند. بنابراین در باب رمزگذاری و رمزگشایی مغز با چالش‌های بیش‌تری روبرو می-شوند. در این پژوهش برای آموزش و تست مدل رمزگشایی پیشنهادی، مجموعه داده‌ای شامل ۱۱/۵ ساعت از داده‌های fMRI، که از ۳ سوژه انسانی جمع‌آوری شده است، مورد استفاده قرار می‌گیرد. در حین اخذ داده‌ها سوژه‌ها در حال تماشای تقریبا ۹۷۲ کلیپ ویدئویی مختلف، شامل اشیاء، صحنه‌ها و کنش‌های متنوع بوده‌اند. این مجموعه داده نسبت به سایر نمونه‌های موجود در مطالعات قبلی، اندازه و پوشش گسترده‌تری دارد [۱۳–۱۵, ۲۸].

۱-۳- اهداف پایان‌نامه

هدف اصلی از انجام این پژوهش، رشد تعامل و پیوند میان دو حوزه علوم اعصاب محاسباتی و هوش‌مصنوعی، به منظور توسعه ماشین‌های هوشمند شبیه به انسان است. با وجود تعامل این دو حوزه و پیشرفت‌های ایجاد شده، هنوز راهی طولانی تا درک کامل مغز بیولوژیکی انسان باقی است.

---

[۱] Adversarial Auto-Encoders

[۲] functional Magnetic Resonance Imaging

دو پرسشی که هنوز در زمینه بینایی پاسخ داده نشده‌اند، عبارت‌اند از:

۱. چگونه مغز انسان اطلاعات بصری را بازنمایی و سازماندهی می‌کند [۲۹].
۲. آیا می‌توان فعالیت‌های مغز را به صورت بلادرنگ رمزگشایی کرد، به گونه‌ای که بتوانیم تفسیری از آنچه که فرد می‌بیند داشته باشیم [۱۶].

در این پژوهش برآنیم تا با ارائه روشی براساس شبکه‌های عصبی، پرسش دوم مطرح شده را تا حد توان پاسخگو باشیم.

۱-۴- ساختار پایان‌نامه

تاکنون در این فصل، مقدمه و کلیات پایان‌نامه، مورد اشاره قرار گرفت و مشخص شد چه مسئله‌ای قرار است حل شود. همچنین اهداف و انگیزه انجام این پژوهش مورد اشاره قرار گرفت. در ادامه، در دومین فصل از گزارش، به معرفی تئوری‌های لازم جهت دنبال کردن روند این پژوهش، می‌پردازیم. به‌گونه‌ای که پس از اتمام این فصل، آشنایی مختصری با هر کدام از مباحث به‌کار گرفته شده، حاصل شود. در فصل سوم ابتدا معرفی کلی از رمزگذاری و رمزگشایی داده‌های fMRI، ارائه می‌شود، و سپس مروری بر روش‌ها و مطالعات موجود در این دو زمینه و ترکیبی از آن‌ها، خواهیم داشت. در ادامه مقایسه‌ای بین شبکه‌های عصبی عمیق و جریان‌های بینایی انسان، از نظر معماری و قوانین محاسباتی حاکم بر آن‌ها خواهیم داشت، که این قیاس به‌نوعی بیانگر ارتباط میان آن‌ها است. در نهایت نشان می‌دهیم که اخیرا مدل‌های مبتنی‌بر شبکه‌های عصبی عمیق، توانسته‌اند نتایج امیدوار کننده‌ای را در مطالعات مربوط به رمزگذاری و رمزگشایی مغز بدست آورند. در ادامه مزایای یک استراتژی یادگیری دوگانه را مورد بحث قرار می‌دهیم. در فصل چهارم روش پیشنهادی پژوهش ارائه می‌شود. به‌گونه‌ای که در هر بخش از این فصل یک گام از گام‌های تشکیل‌دهنده این روش، شرح داده خواهد شد. در فصل پنجم ماحصل آن‌چه که در طی این پژوهش صورت گرفته است و یافته‌های پژوهش مطرح می‌شوند. در آخرین فصل از این گزارش نیز نتیجه‌گیری نهایی انجام می‌گیرد و پیشنهادهایی در جهت توسعه این زمینه موضوعی مطرح می‌شوند.

۲- فصل دوم

مباحث نظری

۲-۱- مقدمه

تصویربرداری پزشکی[۱] تکنیکی است که در جهت ایجاد تصاویری از بدن انسان - بخش‌ها و یا عملکردهای آن - برای اهداف کلینیکی؛ شامل شناخت، درمان و بررسی بیماری‌ها، یا علوم پزشکی؛ شامل مطالعات آناتومیک و فیزیولوژیک، مورد استفاده قرار می‌گیرد. تصویربرداری مغز نقش موثری در پیش‌بُرد مطالعات علوم اعصاب محاسباتی ایفا کرده است. شاخه‌ای از علم که در آن از دیدگاه محاسباتی، به بررسی و درک فرآیندهای موجود در مغز موجودات زنده از جمله انسان پرداخته می‌شود. طی سال‌های اخیر مدل‌های محاسباتی مختلفی برای انجام برخی از وظایف شناختی انسان‌ها ارائه شده‌است که کارایی آن‌ها در برخی از موارد از انسان‌ها نیز بهتر بوده‌است. این اتفاق با ظهور تکنیک‌های جدید در هوش‌مصنوعی، از جمله یادگیری عمیق، سرعت بیشتری به خود گرفته است.

از این‌رو برآنیم در فصل پیش‌ِرو ابتدا انواع تصویربرداری‌های مغزی و به دنبالِ آن، تصویربرداری رزونانس مغناطیسی عملکردی را به‌صورت مختصر معرفی کنیم. سپس مروری بر نحوه تحلیل این دسته از داده‌ها خواهیم داشت. همچنین لازم است مروری بر یادگیری عمیق، شبکه‌های خودرمزگذار و انواع آن‌ها داشته باشیم. در این پژوهش انواع جدیدی از این شبکه‌ها تحت عنوان شبکه‌های خودرمزگذار رقابتی مورد استفاده قرار گرفته‌اند، از این جهت معرفی مفصل‌تری در این زمینه خواهیم داشت.

۲-۲- تصویربرداری‌ساختاری و تصویربرداری‌عملکردی

تصویربرداری مغزی را می‌توان به دو دسته تصویربرداری‌ساختاری[۲] و تصویربرداری-عملکردی[۳] تقسیم‌بندی کرد. در مطالعات انجام شده براساس تصویربرداری‌های ساختاری، هدف،

[۱] Medical imaging
[۲] Structural brain imaging
[۳] Functional brain imaging

تحقیق و بررسی ساختار مغز و تشخیص بیماری‌ها بر اساس آسیب‌های وارده بر ساختار مغز است. از جمله تصویربرداری‌های ساختاری عبارت‌اند از: تصویربرداری رزونانس مغناطیسی (MRI)[۱]، توموگرافی (CAT)[۲] و توموگرافی انتشار پوزیترون (PET)[۳]. این درحالی است که از تصویربرداری عملکردی غالبا به منظور مطالعه پروسه‌های شناختی[۴] استفاده می‌شود. درواقع این دسته از تصویربرداری‌هایِ مغز، در پی یافتن روش‌هایی شکل گرفته‌اند که بتوان به کمک آن‌ها عملکرد مغز را نشان داد. نمونه‌هایی از این دسته، شامل: fMRI و PET هستند. تغییرات عملکردی مغز با استفاده از روش‌های الکتروانسفالوگرافی (EEG)[۵] و مگنتوانسفالوگرافی (MEG)[۶] نیز قابل اندازه‌گیری و بررسی هستند، روش‌هایی که در آن‌ها تغییرات عملکردی مغز، برپایه سیگنال‌های اخذ شده، ثبت می‌شوند. هر کدام از روش‌های عملکردی مزیت‌ها و معایبی از لحاظ رزولوشن زمانی، مکانی و هجوم پذیری[۷] دارند، که براین اساس در مطالعات مختلف ممکن است یک یا چند روش مورد استفاده قرار گیرد.

در حوزه بررسی عملکردی مغز، غالبا روش کار به این صورت است که، سوژه در معرض یک تحریک خارجی هم‌چون بینایی و یا شنوایی قرار می‌گیرد. در حین انجام آزمایش، سیگنال‌ها و تصاویر مورد نیاز از روش‌های نام‌بُرده جمع‌آوری می‌شوند. پس از پردازش و تجزیه‌وتحلیل روی سیگنال‌ها و تصاویر اخذ شده، نقاط فعال در ارتباط با تحریک اعمال شده به دست می‌آیند.

فعالیت مغز به صورت فعالیت توده‌هایی از نرون‌ها است. فعالیت هر نرون در واقع به صورت حرکت یون‌های باردار در غشاء سلول‌های عصبی است. حرکت این یون‌ها در فاصله کوچک غشا

---

[۱] Magnetic Resonance Imaging
[۲] Computed Axial Tomography
[۳] Positron Emission Ttomography
[۴] Cognitive process
[۵] Electroencephalography
[۶] Magnetoencephalography
[۷] Invasiveness

سلولی، مانند یک دوقطبی جریان[۱] کوچک عمل می‌کند و باعث ایجاد پتانسیل الکتریکی روی پوست سر (در روش EEG) و میدان مغناطیسی دور سر (در روش MEG) می‌شود. از آن‌جا که استخوان جمجمه رسانای الکتریکی خوبی نیست، سیگنال‌های EEG تضعیف زیادی دارند. در عوض از نظر مغناطیسی بافت‌های سر دارای خواص یکسانی هستند و از این رو اعوجاج و تضعیف سیگنال MEG به میزان کمتری است. اما قابل بیان است که، سیگنال مغناطیسی مغز حدودا ده میلیون بار از سیگنال مغناطیسی دائمی زمین کوچک‌تر است، بنابراین به‌منظور آشکارسازی این سیگنال‌ها نیاز به تجهیزات خاص نظیر دستگاه‌های ابررسانا کوانتومی تداخلی (SQUID)[۲] است. EEG روشی ارزان و غیرتهاجمی است ولی اشکال اصلی این روش رزولوشن مکانی بسیار پایین آن است. در روش PET، جریان خون و برخی متابولیسم‌های مربوط به آن با اندازه‌گیری اشعه‌ی گامای ساطع شده از یک ماده حاجب که به سیستم گردش خون تزریق شده است، بررسی و ثبت می‌شوند. این روش تهاجمی و بسیار گران بوده و همچنین رزولوشن زمانی آن نیز بسیار پایین است. مزیت این روش امکان استفاده از آن برای قسمت‌های مختلف بدن است. سیگنال‌های EEG و MEG مستقیما فعالیت مغز را نشان می‌دهند، در حالی که در روش‌های fMRI و PET تصویر به دست آمده، حاصل از تغییر یک عامل واسطه است [۳۰].

۲-۳- تصویربرداری رزونانس مغناطیسی عملکردی

می‌توان گفت، fMRI تعادل مناسبی بین رزولوشن زمانی، مکانی و هجوم پذیری برقرار می‌کند و به همین دلیل است که در دهه گذشته به عنوان روش غالب در تصویربرداری عملکردی و پیش‌برد مطالعات فرایندهای شناختی، استفاده می‌شود.

با فعال شدن سلول‌های عصبی در حالت‌های مختلف مانند مطالعه، خوشحالی، عصبانیت، رانندگی، ورزش و غیره، حجم خون ارسالی به قسمت‌های مرتبطِ بدن افزایش می‌یابد. روش، fMRI

---

[۱] Current dipole

[۲] Superconducting Quantum Interface Device

تکنیکی برای مطالعه فعالیت‌های مغز با بررسی تغییرات موضعی متابولیک و همودینامیک آن، یعنی تغییرات جریان و اکسیژن خون، است. دستورهای مغز به غدد درون‌ریز برای ترشح هورمون‌ها، یا اندام‌های مختلف برای انجام حرکات و واکنش‌ها، ناشی از تابع پاسخ همودینامیکی [۱](HRF) است که با توجه به تحریک خارجی ایجاد شده است [۳۰].

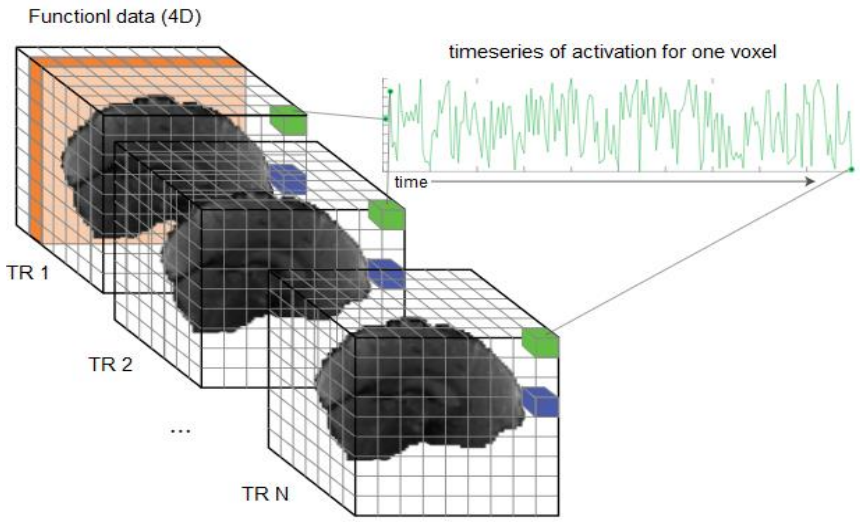
مکانیزم کلی مورد استفاده توسط روش‌های مختلف fMRI به ترتیب زیر است [۳۰]:

۱. دریافت تحریک ورودی
۲. تغییرات منطقه‌ای در سلول‌های عصبی
۳. تغییرات در گردش خون (حجم، جریان، میزان اکسیژن و غیره)
۴. فعال‌شدن نواحی خاص مغز با ایجاد یک کنتراست درونی

حین انجام تصویربرداری fMRI در فواصل زمانی مشخصی، تصاویر سه بعدی آناتومیک سطح خاکستری از مغز اخذ می‌شود؛ که این تصاویر با انتقال داده‌های جمع‌آوری شده از فضای فوریه[۲] به فضای تصویر جهت سهولت پردازش‌های آتی، تشکیل می‌شوند. هر تصویر ایجاد شده شامل وکسل‌های متعددی است. در نهایت می‌توان با دنبال کردن سیگنال وکسل مورد نظر در طول زمان، میزان همبستگی آن را با زمان‌بندی آزمایش طراحی شده به‌دست آورد و در مورد فعال یا غیر فعال بودن آن نظر داد (شکل ۲-۱).

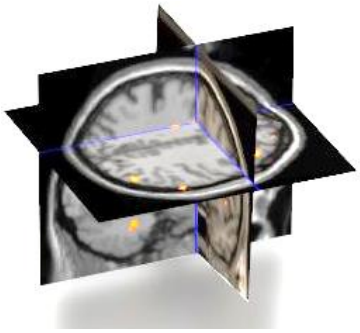
[۱] Hemodynamic Response Function

[۲] k-space



شکل ۲-۱: تصویربرداری در طول زمان براساس روش رزونانس مغناطیسی عملکردی. سری زمانی به ازای هر وکسل خاص، با دنبال کردن مقدار همان وکسل در تصاویری سه بُعدی متوالی حاصل می‌شود [۳۱].

خروجی نهایی روش fMRI، تصاویری است که نواحی رنگی در آن، بیانگر ناحیه‌هایی در مغز هستند که فعالیت انجام شده توسط انسان، به وظایف آن ناحیه‌ها مرتبط است. شکل ۲-۲ یک تصویر سه بعدی (یا سه تصویر دوبعدی، معروف به تصاویر سطح مقطعی) به عنوان پاسخ ناشی از تحریک ورودی را نشان می‌دهد.



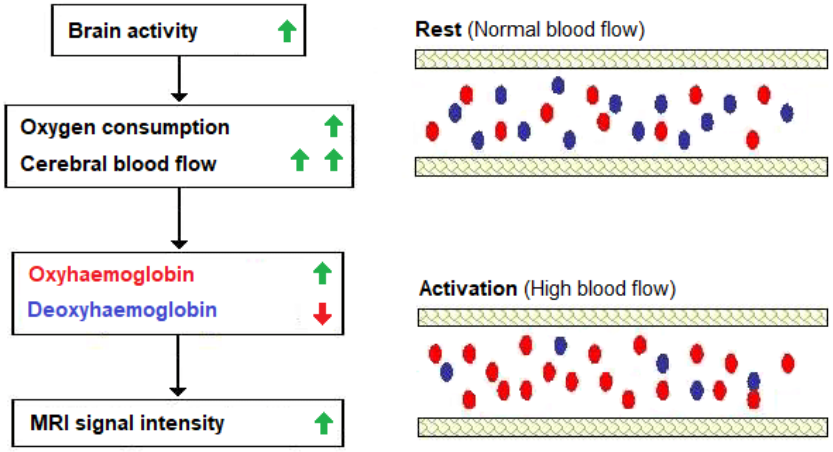
شکل ۲-۲: تصویر خروجی در روش تصویربرداری رزونانس مغناطیسی عملکردی.

از میان روش‌های مطرح در fMRI، روش وابسته به میزان اکسیژن خون (BOLD)[۱] رایج‌ترین راه‌کار است. این روش اولین بار توسط Ogawa مطرح شد [۳۲] و همچنان به عنوان

[۱] Blood Oxygenation level Dependent

متداول‌ترین روش مورد استفاده در fMRI مطرح است که در آن نسبت هموگلوبین اکسیژن‌دار[۱] به هموگلوبین بدون اکسیژن[۲] سنجیده می‌شود.

هموگلوبین اکسیژن‌دار دارای خاصیت دیامغناطیسی و هموگلوبین بدون اکسیژن دارای خاصیت پارامغناطیسی است. بنابراین تغییرات سطح اکسیژن خون باعث تغییرات در شدت سیگنال MRI در تصاویر نهایی خواهد شد [۳۳]. مراحلی که به‌ترتیب در این روش باعث شکل‌گیری سیگنال HRF به ازای هر وکسل می‌شوند، در شکل ۲-۳ نمایش داده شده است.

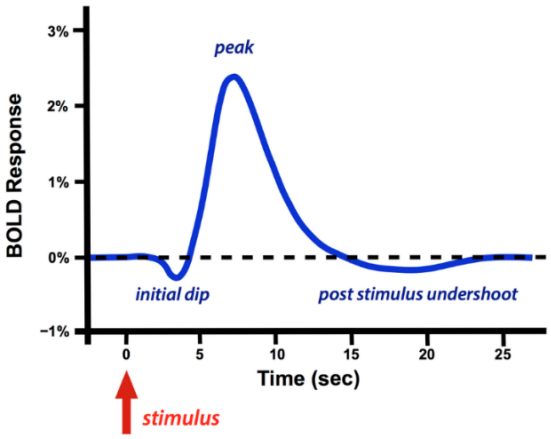


شکل۲-۳: مکانیزم عملکرد روش وابسته به میزان اکسیژن خون [۳۳].

طی یک فعالیت مغزی یا به عبارتی افزایش متابولیسم بدن، مصرف اکسیژن در بافت مورد نظر تقریبا ۵ درصد افزایش می‌یابد، که این امر باعث کاهش فشار اکسیژن می‌گردد. به مدت کوتاهی میزان خاصیت پارامغناطیس افزایش یافته و به دنبال آن از شدت سیگنال HRF کاسته می‌شود. به جبران افزایش مصرف اکسیژن، جریان خون مغز نیز افزایش می‌یابد. بنابراین با افزایش جریان خون، اکسیژن رسانی نیز افزایش می‌یابد، این افزایش در شدت سیگنال HRF نیز مشاهده می‌شود. سپس

[۱] Oxyhemoglobin
[۲] Deoxyhemoglobin

مجددا تعادل هموگلوبین‌ها به حالت استراحت باز می‌گردد. نتیجه چنین مکانیزمی، سیگنالی به فُرم شکل ۲-۴ خواهد بود.



شکل ۲-۴: تابع پاسخ همودینامیکی در روش وابسته به میزان اکسیژن خون [۳۴].

اسکن مغز توسط دستگاه fMRI با رزولوشن پایین‌تری در مقایسه با تصاویر MRI و هر ۲ الی ۴ ثانیه یک بار انجام می‌شود. البته تغییرات سیگنال BOLD در این حالت، بسیار ناچیز است. به عنوان مثال، میزان تغییرات برای دستگاه ۱/۵ تسلا بین ۱ تا ۵ درصد است [۳۵].

چون تغییرات ایجاد شده در میدان مغناطیسی توسط یک نرون فعال، تغییرات ضعیفی است و دارای نسبت کنتراست به نویز (CNR)[۱] کوچک‌تر از یک است، این تغییرات تنها زمانی نمایان می‌شود که بتوان اختلاف میان تصاویر بین زمان‌های استراحت و فعالیت مغز را مشخص نمود.

تصویربرداری fMRI در یک محدوده زمانی چند دقیقه‌ای انجام می‌شود و در این بازه حتما باید اتفاق خاصی بیفتد به‌عنوان مثال، تحریک یا تفاوتی در سری زمانی ایجاد شود، تا در مقایسه تصاویر مجاور با یکدیگر قابل ردیابی و مطالعه باشد.

[۱] Contrast to Noise Ratio

شاید بتوان گفت، جدی‌ترین ایراد وارده به روش BOLD، کیفی بودن بیان اطلاعات در آن باشد. البته تعمیم‌های مختلفی از آن برای حل مشکل و ارائه خروجی به صورت کمی ارائه شده است [۳۶, ۳۷].

همچنین باید این نکته بسیار مهم را متذکر شد که سیگنال‌های BOLD یک اندازه‌گیری غیرمستقیم از فعالیت‌های عصبی هستند و ممکن است به وسیله فعالیت‌های غیرعصبی نیز تحت تاثیر قرار بگیرند و نتیجه نادرستی بدهند [۳۸]. در طراحی آزمایش fMRI باید به این مسئله دقت کافی مبذول داشت و تمامی عوامل محیطی را مد نظر قرار داد.

۲-۴- تحلیل داده‌های تصویربرداری رزونانس مغناطیسی عملکردی

۲-۴-۱- پیش‌پردازش

مرحله پیش‌پردازش به‌منظور آسان‌تر شدن آشکارسازی نواحی فعال، روی داده ها اعمال می‌شود. این مرحله که پیش از تحلیل آماری انجام می‌شود، شامل گام‌هایی است که مستقل از یکدیگر اعمال شده و هر کدام مزیت‌های خاص خود را دارند. همچنین مدت زمان ویژه‌ای لازم دارند. این گام‌ها عبارتند از: اصلاح حرکت سر[۱]، تصحیح زمان‌بندی برش‌های مقطعی[۲] و هموار کردن[۳] دادها با هدف بهبود نسبت سیگنال به نویز (SNR)[۴]. تئوری، کاربرد و نحوه اعمال هر یک از این مراحل در ادامه به‌طور مختصر بیان خواهد شد.

۲-۴-۱-۱- اصلاح حرکت سر

حرکت‌های پیش‌بینی نشده سر سوژه در هنگام برداشت تصویر یکی از منابع ایجاد اغتشاش در داده‌های fMRI است. با وجود اینکه اسفنج‌هایی به‌منظور جلوگیری از این حرکت‌ها در کنار سر سوژه قرار می‌گیرد، اما سر افراد به علت تنفس و عوامل غیرارادی حرکتهای جزئی دارد. تغییرات

[۱] Head motion correction

[۲] Slice timing correction

[۳] Smoothing

[۴] Signal to Noise Ratio

شدت پیکسل‌های حاشیه مغز که بر اثر حرکت ایجاد می‌شود، می‌تواند چندین برابرِ تغییر شدت روشنایی تصویر، ناشی از اثر سیگنال وابسته به سطح اکسیژن خون باشد. همچنین وجود حرکت سر باعث کاهش حساسیت در یافتن مناطق فعال و جابه‌جا شدن برش‌های تصویربرداری می‌شود.

هدف مهم نخستین مرحله پیش‌پردازش، این است که موقعیت مغز برای تمامی تصاویر اخذ شده از یک سوژه، کاملا یکسان باشد. به عبارت دیگر، انجام این مرحله تضمین می‌کند که تفاوت سیگنال مشاهده شده در یک وکسل دلخواه، طی برش‌های مختلف، ناشی از مقایسه میان نواحی کاملا یکسانی از مغز خواهد بود. بنابراین اصلاحاتی روی داده‌های fMRI به‌منظور اصلاح حرکت سر، صورت می‌گیرد. اغلب چنین اصلاحاتی روی هر مقطع به طور جداگانه اعمال می‌شود. به این صورت که اولین تصویر به‌عنوان مرجع در نظر گرفته می‌شود و بقیه تصاویر با آن مقایسه می‌شوند. سپس با اعمال چرخش و انتقال و تغییر اندازه سعی می‌شود مقدار مجموع مربعات اختلاف میان دو تصویر حداقل شود.

این مرحله از پیش‌پردازش، خود تشکیل شده از دو گام مجزا است:

۱. تثبیت[۱]
۲. تبدیل[۲]

از آن‌جا که تغییرات جسم صلب به صورت یکپارچه اعمال می‌شود و با حرکت آن، تمامی نواحی دچار تغییرات مشابه می‌شوند، در مرحله تثبیت، به تخمین پارامترهایی که حرکت سر را به‌عنوان یک جسم صلب توصیف می‌کنند، می‌پردازیم تا بتوانیم تغییرات حرکتی پیش‌آمده را تصحیح کنیم. تغییرات تصاویر، شامل سه درجه آزادی برای بیان انتقال در راستای محورهای X، Y و Z و سه درجه آزادی برای تغییرات چرخش حول همین محورها است.

[۱] Registration
[۲] Transformation

بدیهی است که تغییرات لازمه برای ترازشدن، معکوس تغییرات ناشی از حرکت سر سوژه است. اگر نقطه (x_0, y_0, z_0) به‌عنوان ورودی و (X_0, Y_0, Z_0) به‌عنوان خروجی مرحله تثبیت تعریف شوند و ماتریس M بیانگر تغییرات حرکتی جسم صلب باشد، خواهیم داشت:

$$\begin{aligned} X_0 &= m_{11}x_0 + m_{12}y_0 + m_{13}z_0 \\ Y_0 &= m_{21}x_0 + m_{22}y_0 + m_{23}z_0 \\ Z_0 &= m_{31}x_0 + m_{32}y_0 + m_{33}z_0 \end{aligned} \quad (۱-۲)$$

$$M_{4\times4} = T_{4\times4}R_{4\times4} \quad (۲-۲)$$

ماتریس T تغییرات مربوط به شیفت در راستای سه محور را نشان می‌دهد و تغییرات دورانی در راستای محورها نیز با ماتریس R بیان می‌شود:

$$T = \begin{bmatrix} 1 & 0 & 0 & X_{translation} \\ 0 & 1 & 0 & Y_{translation} \\ 0 & 0 & 1 & Z_{translation} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (۳-۲)$$

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\varphi) & \sin(\varphi) & 0 \\ 0 & -\sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos(\Omega) & \sin(\Omega) & 0 & 0 \\ -\sin(\Omega) & \cos(\Omega) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

(۴-۲)

در دومین گام از اصلاح حرکت سر، بازسازی تصویر با استفاده از درجات مختلف درونیابی (نزدیکترین همسایه، خطی و غیره) انجام می‌شود [۳۰].

۲-۴-۱-۲- تصحیح زمان‌بندی برش‌های مقطعی

در تصویربرداری fMRI، معمولا سر به صورت بُرش‌بُرش تصویربرداری می‌شود. ثبت هر برش مدت زمانی به طول می‌انجامد. بنابراین زمان تصویربرداری از اولین برش از ابتدای سر، تا آخرین برش از انتهای سر، متفاوت است و دارای اختلاف زمانی هستند. درحالی که در زمان آنالیز داده‌ها

فرض بر آن است که از تمامی سر در یک لحظه تصویربرداری شده است. از آنجا که می‌دانیم چنین هم‌زمانی وجود ندارد، داده را به‌گونه‌ای انتقالِ زمانی می‌دهیم که انگار برش اول و آخر در یک زمان تصویربرداری شدند.

این کار زمانی که بخواهیم تمامی وکسل‌ها را در کنار یکدیگر داشته باشیم، لازم است. در غیر این صورت پردازش دچار خطا می‌شود. به عنوان مثال در صورت عدم اصلاح زمان برش، در تشخیص مناطق فعال گمان می‌شود نقطه دوم بعد از نقطه اول فعال شده در نتیجه خطا رخ می‌دهد. به همین منظور اعمال این مرحله الزام می‌یابد.

۲-۴-۱-۳- هموار کردن

هموارسازی به عنوان یک مرحله پیش‌پردازش با هدف از بین بردن نویز انجام می‌گیرد. از آن‌جا که سیگنال وابسته به سطح اکسیژن خون با تغییرات آهسته جریان خون تغییر می‌کند، سرعت تغییرات محدودی دارد. بنابراین هموار کردن سری زمانی پیکسل‌ها میزان نسبت سیگنال به نویز را افزایش می‌دهد. انتخاب فرکانس قطع سیستم پایین‌گذر بستگی به عوامل مختلفی دارد مانند زمان تکرار و دوره تناوب انجام تحریک. در واقع هدف از اعمال این مرحله آن است که، نویزهای فرکانس پایین و بالا را حذف کنیم. در نتیجه‌ی این مرحله، نویز میانی که به‌عبارتی نتیجه فعالیت مغز است، باقی می‌ماند. رایجترین نوع هموارسازی زمانی، استفاده از فیلتر گاوسی است که شکل یک توزیع نرمال دارد. فرم ریاضی این فیلتر به‌وسیله رابطه زیر تعریف می‌شود:

$$f\left(t\right)=e^{-\frac{t^2}{2\omega^2}} \qquad (۵-۲)$$

ω پهنای فیلتر گوسی است. آزمایش‌ها نشان داده است که در استفاده از فیلترهای زمانی نیاز به دقت زیادی است. به‌خصوص از نظر اینکه با هموارکردن زمانی، استقلال داده‌ها و پیکسل‌های مجاور از بین رفته و ممکن است استدلال آماری بر روی داده‌ها دچار مشکل شود.

۲-۴-۲- تحلیل آماری

به‌منظور تحلیل آماری داده‌های fMRI روش‌های مختلفی موجود است. هدف از چنین تحلیل‌هایی ایجاد تصویری است که نواحی فعال شده مغز را به ازای یک تحریک خاص، نشان می‌دهد. حساسیت یافتن مناطق فعال، وابستگی زیادی به انتخاب روش تحلیل داده‌ها دارد. در حالی که انتخاب روش تحلیل نیز به نوع داده‌ها دابسته است [۳۹].

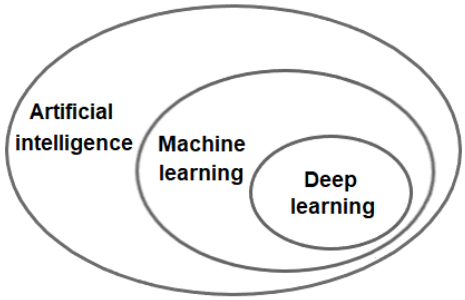
روش‌های تحلیل داده‌های fMRI را می‌توان به دو دسته کلی تقسیم‌بندی کرد:

۱. روش‌های مدل‌محور: این دسته از روش‌ها، با درنظرگرفتن یک مدل اولیه برای شبیه‌سازی تحریک، سیگنال‌های fMRI را بررسی می‌کنند. در اینگونه روش‌ها، حداکثر شباهت میان مدل پیش‌فرض و تصاویر fMRI با پارامترها و تست‌های آماری تعیین می‌شود تا با اعمال حد آستانه‌های مختلف، نواحی فعال در تصاویر مشخص شوند.

۲. روش‌های داده‌محور: در دیگر سو روش‌های داده‌محور قرار دارند که بدون پیش‌فرض در مورد داده‌ها عمل کرده و با استفاده از خود مجموعه داده، به یافتن مشخصه‌ها و کلاس‌بندی و قطعه‌بندی نواحی فعال می‌پردازند. این دسته از روش‌ها برخلاف روش‌های دسته اول، مدل مشخصی را به عنوان فرض اولیه به داده‌ها اعمال نمی‌کنند. تکنیک‌های آنالیز چندمتغیره (مکانی یا زمانی) استفاده‌شده در این روش‌ها، به دنبال یافتن الگوهای مشخصی در داده‌ها هستند که مبین بیشترین تغییرات (واریانس، کوواریانس، انرژی و غیره) در مجموعه وکسل‌ها باشد.

۲-۵- یادگیری عمیق

یادگیری‌عمیق زیرشاخه‌ای از یادگیری‌ماشین و هر دو زیز مجموعه‌ای از هوش‌مصنوعی هستند (شکل ۲-۵).



شکل ۲-۵: رابطه میان یادگیری عمیق و هوش‌مصنوعی. یادگیری‌عمیق زیرشاخه‌ی از یادگیری‌ماشین و هر دو زیر مجموعه‌ای از هوش‌مصنوعی هستند.

یکی از اهداف یادگیری‌عمیق، جایگزین کردن ویژگی‌های دستی با الگوریتم‌های کارآمد برای یادگیری بدون‌سرپرست[۱] و یا شبه‌سرپرست[۲] و استخراج ویژگی‌ها به صورت سلسله مراتبی است.

شبکه‌های عصبی عمیق عملکرد فوق‌العاده‌ای در کاربردهای بینایی بسیار سخت، مانند رقابت ImageNet، به نمایش گذاشته‌اند [۴۰]. به‌طوری‌که اکنون این شبکه‌ها جز موفق‌ترین الگوریتم‌های استفاده شده محسوب می‌شوند. عمده روش‌های موفق در یادگیری عمیق، مانند شبکه-های همگشتی عمیق (CNNs)[۳]، خودرمزگذارها (AEs)[۴] و شبکه‌های باور عمیق (DBNs)[۵] همگی بر پایه شبکه‌های عصبی مصنوعی بنا نهاده شده‌اند.

۲-۶- شبکه‌های خودرمزگذار

خودرمزگذارها خانواده‌ای از شبکه‌های عصبی با یادگیری بدون سرپرست یعنی مسایلی که در آن برچسبی برای توصیف داده‌ها وجود ندارد، هستند. هدف از به‌کارگیری این شبکه‌ها تشخیص الگوهای ذاتی در یک مجموعه داده است. همچنین می توانند برای برچسب زدن به الگوهای بدست آمده، استفاده شوند. در اصل، خودرمزگذارها مجموعه‌ای از داده‌های بدون برچسب را دریافت

[۱] Unsupervised
[۲] Semisupervised
[۳] Convolutional Neural Networks
[۴] Auto-Encoders
[۵] Deep Belief Networks

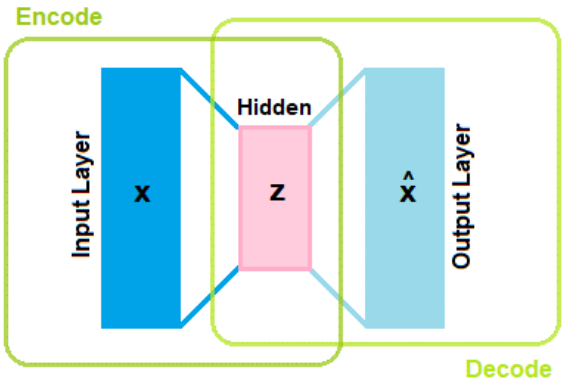
می‌کنند و در خروجی سعی بر بازنمایی داده‌های ورودی دارند، به‌گونه‌ای که کمترین اختلاف ممکن را با مقادیر ورودی داشته باشند. در حین این فرآیند، ساختار ذاتی داده‌ها را کشف کرده و ویژگی‌های مهم آن‌ها را استخراج می‌کنند. استخراج ویژگی‌ها به صورت خودکار مزیت این شبکه‌ها محسوب می‌شود. بنابراین می‌توان گفت ساختار خودرمزگذارها به دو بخش رمزگذاری و رمزگشایی تقسیم می‌شود. طی بازنمایی انجام شده توسط این شبکه به ازای هر نمونه از مجموعه داده $\left\{x^{(1)},...,x^{(t)}\right\}$ داریم:

$$\begin{aligned} z^{(t)} &= f_\theta\left(x^{(t)}\right) \\ \hat{x}^{(t)} &= g_\theta\left(z^{(t)}\right) \end{aligned} \qquad (۶-۲)$$

f_θ و g_θ به ترتیب تحت عنوان توابع رمزگذار و رمزگشا شناخته می‌شوند. در مرحله رمزگذاری داده ورودی، $x^{(t)}$، به فُرم بردار ویژگی، $z^{(t)}$، تبدیل می‌شود. سپس در مرحله رمزگشایی از روی ویژگی‌های حاصل شده، بازنمایی از داده‌های ورودی، $\hat{x}^{(t)}$، خواهیم داشت.

در حالت معمول خودرمزگذارها شبکه‌های کم عمقی هستند که رایج‌ترین آن‌ها دارای یک لایه ورودی، یک لایه نهان و یک لایه خروجی است (شکل ۲-۶). این مفاهیم برای اولین بار، در سال ۱۹۸۰ توسط Hinton و گروه تحقیقاتی PDP مطرح شد [۴۱] . این در حالی است که، مجددا در دهه‌ی ابتدایی قرن بیست و یکم در معماری عمیق به فرم ماشین بولتزمن محدود شده (RBM)[۱] مورد توجه قرار گرفتند. در خودرمزگذارها با ساختارهای عمیق‌تر، شبکه، تعدادی برابر از لایه‌ها را برای رمزگذاری و رمزگشایی در اختیار دارد. کاربرد اصلی ساختارهای عمیق، در زمینه کاهش ابعاد است، که هزینه‌های زمانی و حافظه‌ای پردازش را کاهش می‌دهند.

[۱] Restricted Boltzmann Machine



شکل ۲-۶: معماری کلی شبکه‌های خودرمزگذار. ساختار متشکل از دو بخش، رمزگذاری و رمزگشایی است. معماری این شبکه‌ها شامل لایه ورودی، لایه نهان و لایه خروجی است.

شبکه‌های خودرمزگذار انواع الگوریتم‌های مختلفی را شامل می‌شوند:

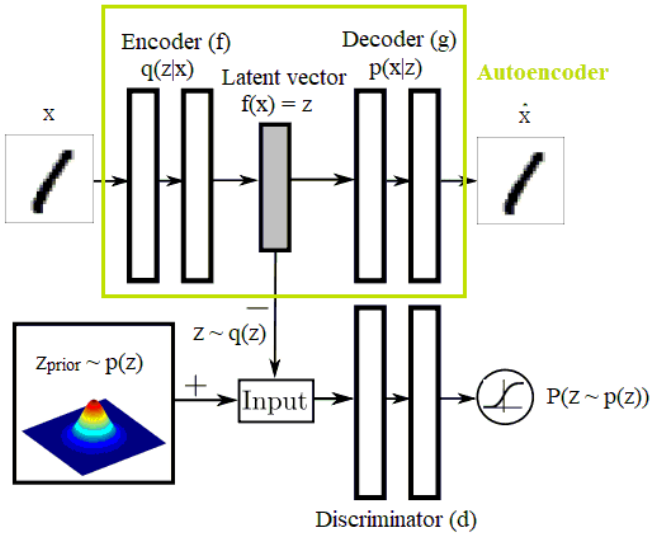
- خودرمزگذار پشته‌ای (SAE)[۱]
- خودرمزگذار حذف‌نویز (DAE)[۲]
- خودرمزگذار تُنُک (SAE)[۳]
- خودرمزگذار همگشتی (CAE)[۴]
- خودرمزگذار متغیّر
- خودرمزگذار رقابتی

۲-۷- شبکه‌های خودرمزگذار رقابتی

شبکه‌های خودرمزگذار رقابتی برای اولین بار توسط Makhzani [۴۲] مطرح شده‌اند که می‌توانند یک شبکه خودرمزگذار را به یک مدل تولیدی تبدیل کنند.

[۱] Stacked Auto-Encoder
[۲] Denoising Auto-Encoder
[۳] Sparse Auto-Encoder
[۴] Convolutional Auto-Encoder

در اصل در مدل ارائه شده، یک شبکه خودرمزگذار با اهداف دوگانه آموزش داده می‌شود؛ یک معیار خطای بازسازی متداول، و یک معیار آموزش رقابتی [۱۹] که توزیع پسین[۱] تجمعی فضای‌نهان خودرمزگذار را با توزیع پیشین[۲] دلخواه مطابقت می‌دهد. این معیار آموزشی ارتباطی قوی با آموزش VAE دارد. نتیجه آموزش این است که رمزگذار یاد می‌گیرد توزیع داده‌ها را به توزیع پیشین تبدیل کند، در حالی که رمزگشا مدل مولدی را می‌آموزد که توزیع پیشین تحمیل شده را به توزیع داده‌ها نگاشت می‌کند. شکل ۲-۷ دیدگاهی اولیه از روند این شبکه در ذهن ایجاد می‌کند.



شکل ۲-۷: معماری یک شبکه خودرمزگذار رقابتی. ردیف بالا یک شبکه خودرمزگذار است که تصویر x را از روی کد نهان z بازسازی می‌کند. نمودار ردیف پایین نیز شبکه دومی را آموزش می‌دهد که به تفکیک، پیش بینی می‌کند که آیا نمونه از کد نهان شبکه خودرمزگذار ناشی می‌شود یا از توزیع مشخص شده توسط کاربر [۴۲].

در نظر بگیرید که x ورودی و z کد نهان (فضای نهان) یک شبکه خودرمزگذار شامل یک زیرشبکه رمزگذار و یک زیرشبکه رمزگشا، باشند. $p(z)$ توزیع پیشین باشد که می‌خواهیم بر روی کدها تحمیل کنیم، $q(z|x)$ و $p(x|z)$ نیز به ترتیب توزیع کدگذاری و توزیع رمزگشایی باشند. همچنین در نظر بگیرید که $p_d(x)$ توزیع داده و $p(x)$ توزیع مدل باشند. در این صورت تابع

[۱] Posterior distribution

[۲] Prior distribution

رمزگذارِ شبکه خودرمزگذار، $q(z|x)$، توزیع پسین تجمعی $q(z)$ را، بر بردار کد نهان شبکه خودرمزگذار، به شرح زیر تعریف می‌کند:

$$q(z)=\int_X q(z|x)\,p_d(x)\,dX \qquad (۲-۳)$$

می‌توان گفت، خودرمزگذار رقابتی یک شبکه خودرمزگذار است که به وسیله تطبیق توزیع پسین تجمعی، $q(z)$، با توزیع پیشین تحمیل‌شده، $p(z)$، تنظیم می‌شود. به منظور انجام این کار، از یک زیرشبکه تفکیک‌کننده، که به کد نهان شبکه خودرمزگذار متصل شده است و در شکل ۲-۷ نیز نشان داده شده است، بهره می‌گیرد. در واقع این زیرشبکه تفکیک‌کننده است که $q(z)$ را به منظور مطابقت با $p(z)$ راهنمایی و هدایت می‌کند. ضمنا در همین حین، شبکه خودرمزگذار سعی بر، به حداقل رساندن خطای بازسازی دارد. مولد شبکه خودرمزگذار رقابتی نیز، رمزگذارِ موجود در شبکه خودرمزگذار است، $q(z|x)$. رمزگذار تضمین می‌کند که توزیع پسین تجمعی می‌تواند شبکه تفکیک‌کننده رقابتی را فریب دهد و فکر کند کد نهان $q(z)$ از توزیع پیشین واقعی $p(z)$ بدست می‌آید.

هر دو، زیرشبکه تفکیک‌کننده و شبکه خودرمزگذار در دو مرحله - مرحله بازسازی و مرحله تنظیم - به طور مشترک با استفاده از روش گرادیان کاهشی تصادفی (SGD)[۱] به ازای چند نمونه[۲] آموزش می‌بینند. در مرحله بازسازی، شبکه خودرمزگذار، زیرشبکه‌های رمزگذار و رمزگشا را به‌روزرسانی می‌کند، به گونه‌ای که خطای بازسازی ورودی‌ها را به حداقل برساند. در مرحله تنظیم، زیرشبکه تفکیک‌کننده ابتدا خود را به‌روزرسانی می‌کند تا نمونه‌های واقعی (تولید شده بر اساس توزیع پیشین) را از نمونه های تولید شده (کدهای نهان محاسبه شده توسط خودرمزگذار) جدا کند.

[۱] Stochastic Gradient Descent

[۲] Mini batch

سپس شبکه خودرمزگذار، رمزگذارِ خود را به‌روزرسانی می‌کند تا زیرشبکه تفکیک‌کننده را به اشتباه بیندازد.

پس از اتمام مراحل آموزش، زیرشبکه رمزگشایِ شبکه خودرمزگذار، یک مدل مولدی را تعریف می‌کند که توزیع پیشین تحمیل شده $p(z)$، را به توزیع داده‌ها نگاشت می‌کند.

چندین گزینه احتمالی برای رمزگذارِ شبکه خودرمزگذار رقابتی، $q(z \mid x)$، وجود دارد:

- مدل قطعی[۱]:

در اینجا فرض می‌کنیم که $q(z \mid x)$ یک تابع معینی از x است. در این حالت، رمزگذار مشابه رمزگذارِ یک خودرمزگذار استاندارد است و تنها منبع تصادفی در $q(z)$، توزیع داده‌ها، یعنی $p_d(x)$، است.

- مدل گاوسی پسین[۲]:

در اینجا فرض می کنیم که $q(z \mid x)$ یک توزیع گاوسی است که میانگین و واریانس آن توسط شبکه رمزگذار پیش‌بینی می‌شود: $z_i \sim \mathcal{N}(\mu_i(x), \sigma_i(x))$. در این حالت، تصادفی بودن $q(z)$، هم از توزیع داده و هم از تصادفی بودن توزیع گاوسی در خروجی رمزگذار ناشی می‌شود. می‌توان از ترفند پارامتری سازی مجدد [۱۷] برای پس انتشار از طریق شبکه رمزگذار استفاده کرد.

- تقریب جامع پسین[۳]:

از شبکه خودرمزگذار رقابتی می توان برای آموزش $q(z \mid x)$ به عنوان تقریبی جامع از توریع پسین استفاده کرد. فرض کنید رمزگذارِ شبکه خودرمزگذار رقابتی تابع $f(x, \eta)$ است که ورودی x و یک نویز تصادفی η را با یک توزیع ثابت (به

[۱] Deterministic

[۲] Gaussian posterior

[۳] Universal approximator posterior

عنوان مثال، گاوسی) می‌گیرد. می‌توان با ارزیابی $f(x,\eta)$ در نمونه های مختلف η، از توزیع پسین دلخواه $q(z|x)$ نمونه گرفت. به عبارت دیگر، می‌توان فرض کرد؛ $q(z|x,\eta)=\delta(z-f(x,\eta))$، آنگاه $q(z|x)$ توزیع پسین و $q(z)$ توزیع پسین تجمعی به شرح زیر تعریف می‌شوند:

$$q(z|x)=\int_{\eta} q(z|x,\eta)\,p_{\eta}(\eta)\,d\eta \qquad (۳-۳)$$

$$q(z)=\int_{x}\int_{\eta} q(z|x,\eta)\,p_{d}(x)\,p_{\eta}(\eta)\,d\eta dx \qquad (۴-۳)$$

در این حالت، تصادفی بودن $q(z)$، هم از توزیع داده و هم از نویز تصادفی η در ورودی رمزگذار ناشی می‌شود. توجه داشته باشید که در این حالت توزیع پسین $q(z|x)$ دیگر محدودیتی برای گاوسی بودن ندارد و رمزگذار می‌تواند هر توزیع پسین دلخواهی را برای ورودی x مشخص کند. از آنجا که روشی کارآمد برای نمونه برداری از توزیع پسین تجمعی $q(z)$ وجود دارد، روش آموزش رقابتی قادر است $q(z)$ را با $p(z)$ را به وسیله پس انتشار از طریق تابع $f(x,\eta)$ زیرشبکه رمزگذار، مطابقت دهد.

انتخاب انواع مختلف $q(z|x)$ منجر به ایجاد انواع مختلف مدل‌ها با پویایی آموزشی متفاوت می‌شود. به عنوان مثال، در مدل قطعی $q(z|x)$، شبکه باید فقط با بهره گیری از تصادفی بودن توزیع داده‌ها، $q(z)$ را با $p(z)$ مطابقت دهد، اما از آنجا که توزیع تجربی داده‌ها توسط مجموعه آموزشی ثابت شده است، و نگاشت معین است، ممکن است، $q(z)$ تولید شود که خیلی نرم و هموار نیست. با این حال، در خصوص مدل‌های گاوسی و یا تقریب جامع، شبکه به منابع تصادفی دیگری دسترسی دارد، که با هموارسازی $q(z)$، قادرند در مرحله تنظیم به شبکه کمک کنند.

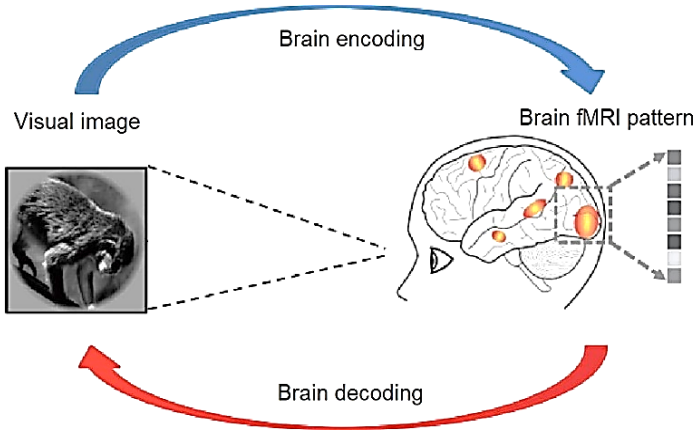
۳- فصل سوم

مروری بر کارهای گذشته

۳-۱- مقدمه

طی دهه گذشته، محققان fMRI، تکنیک‌های رو به رشدی را در جهت تجزیه و تحلیل اطلاعات موجود در فعالیت‌های مغز، توسعه داده‌اند. هدف بسیاری از مطالعات در این راستا، این است که درک کنیم چه اطلاعات حسی، شناختی یا حرکتی در مناطق مختلف مغز ارائه می‌شود. بیشترین میزان درک فعلی، با بهره‌گیری از دیدگاه آینه‌گونه رمزگذاری و رمزگشایی، به‌منظور بررسی داده‌های fMRI حاصل شده است. به شکلی که هنگام تجزیه و تحلیل داده‌ها از منظر رمزگذاری، فرد سعی بر درک این دارد که، در برابر تنوع موجود در جهان، فعالیت مغز و تغییرات آن به چه شکل بوده است. در حالی که هنگام تجزیه و تحلیل داده‌ها از دیدگاه رمزگشایی، تلاش فرد بر تعیین آن است که، با مشاهده‌ی فعالیت‌های مغز، به چه میزان می‌توان در مورد جهان آموخت و آن را درک کرد. در ارتباط با هر دیدگاه تکنیک‌های محاسباتی مختلفی وجود دارد [۱۶].

رمزگذاری و رمزگشایی داده‌های fMRI، دو جنبه مهم در حوزه علوم اعصاب ادراک بصری هستند [۴۳]، و ابزاری مهم جهت شناخت سامانه درک بصری محسوب می‌شوند. یک مدل رمزگذاری بصری، تلاش می‌کند تا پاسخ مغز را براساس محرک بصریِ معین [۴۴, ۴۵] پیش‌بینی کند؛ در حالی که یک مدل رمزگشایی بصری، سعی دارد با تجزیه و تحلیل پاسخ داده شده از سمت مغز، محرک بصری مربوطه را پیش‌بینی کند [۵۸–۴۶, ۲۷, ۱۳, ۱۲]. به این ترتیب از آنجا که رمزگذاری و رمزگشایی مغز (شکل ۳-۱) بینش‌های متعددی در مورد عملکرد مغز ارائه می‌دهند، به دو روش مهم جهت ترویج و پیشرفت حوزه علوم عصبی-حسی تبدیل شده‌اند [۴۳].



شکل ۳-۱: رمزگذاری و رمزگشایی مغز. مدل رمزگذاری تلاش می‌کند تا پاسخ مغز را بر اساس محرک‌های بینایی ارائه شده پیش‌بینی کند، در حالی که مدل رمزگشایی سعی دارد با تجزیه و تحلیل پاسخ‌های مشاهده شده مغز، محرک‌های دیداری مربوطه را استنباط کند. وحدت موثر رویکردهای رمزگذاری و رمزگشایی می تواند پیش‌بینی-های دقیق‌تری را فراهم کند و درک ما از نمایش اطلاعات در مغز انسان را تسهیل کند [۴۳].

در این فصل ابتدا مروری بر روش‌های رمزگذاری و رمزگشایی موجود خواهیم داشت. در ادامه استراتژی‌های دوسویه را مورد بحث قرار می‌دهیم. سپس مقایسه‌ای بین شبکه‌های عصبی عمیق و جریان‌های بینایی انسان، از نظر معماری و قوانین محاسباتی حاکم بر آن‌ها خواهیم داشت، که این قیاس بیانگر ارتباط میان آن‌ها است.

در نهایت نشان می‌دهیم که اخیرا مدل‌های مبتی‌بر شبکه‌های عصبی عمیق، توانسته‌اند نتایج امیدوار کننده‌ای را در مطالعات مربوط به رمزگذاری و رمزگشایی مغز بدست آورند. در آخر مزایای یک استراتژی یادگیری دوگانه را بررسی می‌کنیم.

۳-۲- مدل‌های رمزگذاری

در مطالعات انجام شده تا به امروز، شمار زیادی از مدل‌های رمزگذاری براساس قوانین محاسباتی خاصی ارائه شده‌اند. محققان بر این باورند که این قوانین محاسباتی خاص، می‌توانند مبنایی ریاضی برای پاسخ مغز به محرک‌های بصری باشند. به عنوان مثال Kay و همکارانش [۵۹]، به منظور طراحی یک مدل رمزگذاری، از فیلترهای موجک گابور به شکل هرمی[۱] استفاده کرده‌اند.

[۱] Pyramid-shaped gabor wavelet filters

براساس مدل ارائه شده، با موفقیت می‌توان تصاویر طبیعی که فردِ ناظر مشاهده کرده است را شناسایی کرد. چند سال پس از آن در مطالعه‌ای دیگر Kay و همکارانش [۴۴] ، یک مدل رمزگذاری آبشاری دو مرحله‌ای نیز ارائه داده‌اند. به تازگی st-Yues و Naselaris [۴۵]، مدل ویژگی‌های وزن‌دار را براساس نقشه‌های ویژگی میانی یک شبکه عصبی عمیق از پیش آموزش دیده شده، ارائه داده‌اند. این مدل می‌تواند برای پیش‌بینی پاسخ وکسل‌ها[۲] و مطالعه میدان تاثیر[۳] هر وکسل مورد استفاده قرار گیرد. علاوه‌بر آن، Zeidman و همکارانش [۵۱]، یک مدل جمعیتی بیزین (pRF)[۴]، جهت مطالعات تفسیری کدگذاری مغز مطرح نموده‌اند. چند سال اخیر، شبکه‌های عصبی عمیق موفقیت-های بزرگی در زمینه بینایی ماشین کسب کرده‌اند و محققان استفاده از این شبکه‌ها را برای ارئه مدل‌های پیچیده‌تر رمزگذاری مغز آغاز کرده‌اند [۱۲, ۱۳, ۴۵].

۳-۳- مدل‌های رمزگشایی

مطالعات انجام شده حاکی از امکان رمزگشایی الگوهای کنتراست باینری [۵۵, ۵۷]، نویسه‌های دست‌نویس [۴۶, ۵۸]، تصاویر چهره انسان [۲۰, ۴۷]، تصاویر طبیعی و محرک‌های ویدئویی [۱۲, ۶۰] و رویا [۴۹, ۵۵]، از روی الگوهای فعالیت مغز هستند. در رمزگشایی هدف بازسازی محرک بصری یا به عبارتی تولید تصویری قابل درک از تصویر محرک است. بازسازی تصویر بصری یک مسئله چالش برانگیز است، چرا که نسبت به شناسایی محرک به اطلاعات بیشتری نیازمند است. این نیازمندی در رابطه با بازسازی تصاویر طبیعی که حاوی پیچیدگی و پویایی بالایی هستند، مشهودتر است. به منظور ساده‌تر کردن مسئله بازسازی، اکثر مطالعات بر بازسازی تصاویر ساده متمرکز شده‌اند [۶۱].

---

[۲] Voxels

[۳] Receptive field

[۴] Bayesian population receptive field

به عنوان مثال، Miyawaki و همکاران [۵۵]، یک مدل رمزگشایی چند مقیاسه‌ای به منظور بازسازی الگوهای کنتراست باینری درک شده توسط کاربر از روی پاسخ‌های مغزی، ارائه داده‌اند. مدل ارائه شده، نگاشتی خطی از وکسل‌های قشر بینایی به هر پیکسل تصویر است. در مطالعه‌ای دیگر، Schoenmakers و همکاران [۱۵]، یک مدل گوسین خطی، جهت بازسازی حروف دست‌نویس BRAIN معرفی کرده‌اند. در مدل ارائه شده به منظور بازسازی از داده‌های fMRI قشر بینایی استفاده می‌شود. البته مدتی پس از آن با ارائه یک مدل ترکیبی گوسی، نتایج بازسازی را بهبود بخشیده‌اند [۴۶, ۶۲]. Güçlütürk و همکارانش [۲۰] نیز، ترکیبی از استنتاج احتمالی و آموزش رقابتی را در جهت بازسازی چهره‌های درک شده توسط کاربر، از روی پاسخ‌های مغزی پیشنهاد کرده‌اند. Naselaris و همکاران [۵۳]، برای اولین بار بازسازی تصاویر طبیعی را با استفاده از اطلاعات پیشینی و ترکیبی از مدل‌های کدگذاری ساختاری و کدگذاری معنایی، عملی کرده‌اند. با این حال، مطالعه انجام شده، بیشتر به یک مسئله شناسایی تصویر در یک کتابخانه تصاویر طبیعیِ محدود، شبیه است. به همین نحو، آن‌ها با بازسازی فریم به فریم تصاویر، ویدئو محرک را بازسازی کرده‌اند.

در این میان، شبکه‌های عصبی عمیق، به دلیل قابلیت بالا در بازنمایی ویژگی‌های تصاویر، در سال‌های اخیر مورد توجه محققان قرار گرفته‌اند. این شبکه‌ها پیشرفت قابل ملاحضه‌ای در تشخیص و طبقه‌بندی تصاویر [۶۳, ۶۴]، تشخیص گفتار [۶۵] و غیره دست‌یافته است. بنابراین محققان بیش از پیش، شبکه‌های عصبی عمیق را در مطالعات رمزگشایی بصری داده‌های fMRI، اعمال کرده‌اند [۶۶].

از جمله، Wen و همکاران [۱۲]، یک روش رمزگشایی عصبی پویا را براساس یادگیری عمیق ارائه داده‌اند که قادر است تصاویر بصری پویا درک شده توسط کاربر را بازسازی کرده، و برچسب‌های معنایی آن‌ها را بازنمایی کند. در این مطالعه صحنه‌های یک ویدئو، فریم به فریم بازسازی شدند. Harikawa و همکاران [۶۰]، یک مدل رمزگشایی براساس ویژگی‌های بصری سلسله مراتبی تولید شده توسط DNN، ارائه داده‌اند. یافته‌های آن‌ها نشان می‌دهد که ویژگی‌های بصری

سلسله مراتبی را می‌توان از روی داده‌های fMRI پیش‌بینی کرد و از آن‌ها برای شناسایی دسته‌های اشیا دیده شده، استفاده کرد. علاوه‌بر این آن‌ها دریافته‌اند که، ویژگی‌های رمزگذاری شده براساس داده‌های fMRI، با ویژگی‌های سطح بالا و میانی اجسام تجسم شده در لایه‌های DNN، همبستگی مثبت بالایی دارند [۴۹].

اکثر مطالعات رمزگشایی ذکر شده تا به اینجا، مبتنی بر روش چند وکسل (MVPA)[۵] هستند [۵۲]. با این حال، الگوهای اتصالات مغز[۶] نیز یکی از ویژگی‌های اصلی حالت مغز هستند که، می‌توانند برای رمزگشایی مغز مورد استفاده قرار گیرند. مطالعات پیشین در زمینه رمزگشایی [۶۷– ۷۱] حاکی از آن هستند که می‌توان از اطلاعات اتصال مغز به عنوان ویژگی‌های متمایز کننده در روش‌های رمزگشایی استفاده کرد.

به عنوان مثال؛ Yargholi و Hossein_Zadeh [۷۰] توانسته‌اند با استفاده از اطلاعات اتصالات مغز در رمزگشایی مغز، دو رقم دست‌نویس ۶ و ۹ را با موفقیت بازسازی کنند. Manning و همکارانش نیز [۷۱]، یک مدل احتمالی را برای الگوهای اتصال عملکردی پویا در فعالیت‌های مغز ارائه داده‌اند. این الگوهای اتصالی پیشنهادی را می‌توان در مطالعات رمزگشایی مغز مورد استفاده قرار داد.

۳-۴- رمزگذاری و رمزگشایی ترکیبی با مدل‌های دوسویه

اگرچه تحولات اخیر در رمزگذاری و رمزگشایی مغز [۷۲, ۷۳] بیان‌کننده نتایج امیدوارکننده‌ای بوده‌اند، اما هنوز چالش‌های بسیاری در جهت ساخت یک مدل رمزگشایی دقیق به منظور بازسازی محرک‌های بصری از روی داده‌های fMRI، باقی‌مانده است.

[۵] Multi-Voxel Pattern Analysis

[۶] Brain connectivity patterns

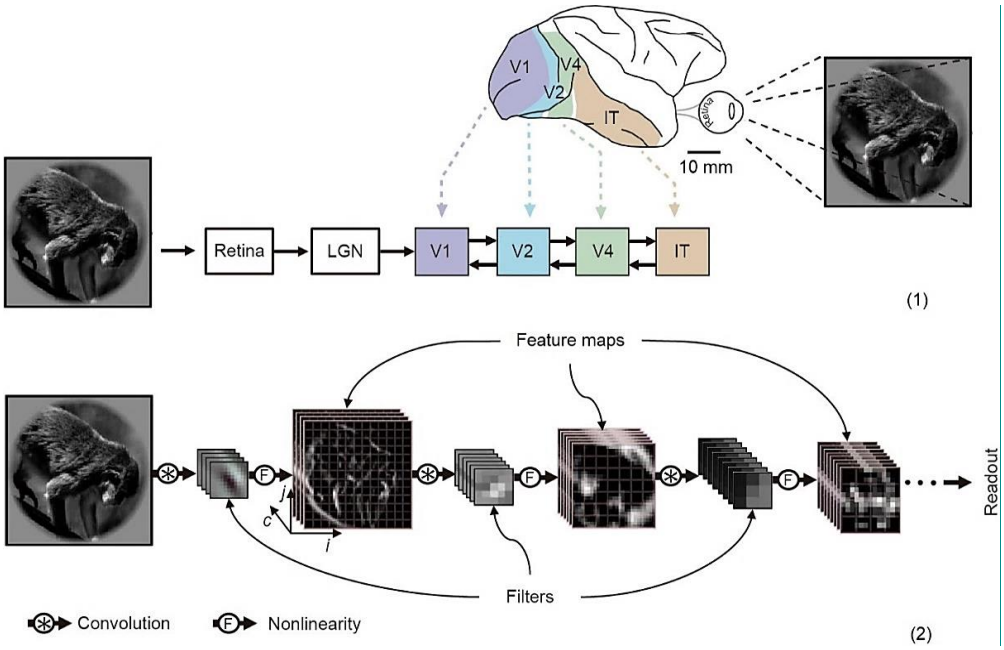
از نقطه‌نظر یادگیری ماشین به زبان قضیه بیز، با استفاده از یک مدل تولیدی که فعالیت مغز را اندازه گیری می‌کند، می‌توان به یک مدل رمزگذاری دست یافت. هنگامی که این مدل رمزگذاری با دانش پیشین در مورد محرک‌ها ترکیب شود، با توجه به الگوی فعالیت مغز، توزیع احتمالی پسین محرک‌ها - یعنی توزیع پیش‌بینی‌کننده برای رمزگشایی - قابل دست‌یابی است. بنابراین، مدل‌های رمزگذاری و رمزگشایی نباید به عنوان یک امر جدا از هم تلقی شوند. یکی کردن موثر رویکردهای رمزگذاری و رمزگشایی می‌تواند پیش‌بینی‌های دقیقی را منجر شود و درک بازنمایی اطلاعات در مغز انسان را تسهیل کند [۵۶, ۷۴].

به عنوان مثال، Fujiwara و همکاران [۵۶]، رویکردی دوسویه به‌منظور بازسازی تصاویر بصری ارائه دادند، که در آن مجموعه‌ای از متغیرهای نهان برای ارتباط پیکسل‌های تصویر و وکسل-های fMRI فرض شده است. این روش امکان برآورد را برای هر دو رویکرد رمزگذاری و رمزگشایی، فراهم می‌سازد. این محققان از چارچوب تحلیل همبستگی متعارف بیزین (BCCA)[۷] استفاده کردند، که چندین تناظر از طریق متغیرهای نهان، بین پیکسل‌های تصویر و وکسل‌های fMRI محاسبه می‌کند. از آنجا که می‌توان درنظر گرفت، وزن‌های پیکسل برای هر متغیر نهان، اساس تصویر را تعریف می‌کند، آموزش مدل تحلیل همبستگی متعارف بیزین با استفاده از داده‌های اندازه‌گیری شده، منجربه استخراج خودکار شالوده تصاویر می‌شود. اگرچه گمانه‌زنی درباره مفاهیم عملکردی شالوده‌های تخمینی تصاویر زودهنگام است، اما این رویکرد دوسویه مبتنی‌بر داده، می‌تواند به‌منظور کشف معماری مدولار مغز در بازنمایی محرک‌های طبیعی پیچیده، رفتار و تجربه‌های ذهنی تعریف شده در فضاها با ابعاد بیشتر، گسترش یابد.

[۷] Bayesian Canonical Correlation Analysis

۳-۵- ارتباط بین شبکه‌های عصبی عمیق و سیستم بینایی انسان

یادگیری عمیق [۹, ۷۵] زیرشاخه بزرگی از روش‌های یادگیری ماشین است که بازنمایی‌هلی سلسله مراتبی را از داده‌های ورودی استخراج می‌کند. معماری شبکه‌های عصبی عمیق برای اولین بار از ساختار و اصول محاسباتی سیستم عصبی بیولوژیکی الهام گرفته است [۷۶]. اخیرا، روش‌های یادگیری عمیق مبتنی‌بر شبکه‌های عصبی عمیق موفقیت‌های بزرگی در زمینه‌های بازشناسی تصویر، تشخیص گفتار، پردازش زبان طبیعی و جنبه‌های دیگر کسب کرده‌اند. از نظر معماری، لایه‌های سلسله مراتبی شبکه‌های عصبی عمیق بسیار شبیه به سیستم بینایی شکمی مغز انسان است [۶, ۹, ۱۲] (شکل ۳-۲). از نظر عملکرد، تحقیقات موجود در زمینه رمزگذاری و رمزگشایی عصبی، بر اساس یادگیری عمیق، نشان داده‌اند که بازنمایی‌های کم عمق شبکه‌های عصبی عمیق، شبیه به عملکرد منطقه بینایی اولیه است، در حالی که بازنمایی‌های عمیق‌تر این شبکه‌ها شبیه به انتهای پشتی سیستم بینایی شکمی است [۱۳, ۶۰, ۷۷, ۷۸].



شکل ۳-۲: سیستم بینایی شکمی و یک شبکه عصبی همگشتی عمیق. (۱) تصویرسازی‌های پیش‌رو و پس‌رو بین چهار ناحیه برودمن (V_1، V_2، V_4 و IT) [۶]؛ (۲) تصویری از یک شبکه عصبی عمیق پیچشی پیش‌رو ساده، که از ساختار سلسله مراتبی آن برای شبیه‌سازی نمایش سلسله مراتبی سیستم بینایی شکمی استفاده می‌شود [۴۵].

انسان‌ها می‌توانند اشیاء پیچیده را به سرعت و با دقت از طریق جریان شکمی بصری درک کنند، سیستمی تشکیل شده از مناطق به هم پیوسته مغز، که ویژگی‌های پیچیده‌ای را در ساختارهای سلسله مراتبی پردازش می‌کند [۸۱–۷۹]. با این حال کشف خودکار مفاهیم بصری از روی تصاویر، بدون وجود اطلاعات نظارت شده، چالشی عمده در تحقیقات یادگیری ماشین است.

از دیدگاه معمول، استفاده از مدل شبکه عصبی عمیق از پیش آموزش دیده، برای یادگیری چنین بازنمایی‌هایی از تصاویر، دشوار است. چرا که معنای هر بعد در بردار بازنمایی که به صورت خودکار از تصویر ورودی توسط مدل شبکه عصبی عمیق استخراج شده است، نامشخص است. بدون ایجاد بازنمایی‌های متعدد و جدا از هم، تفسیر این بازنمایی‌ها در میان وظایف مختلف دشوار است. خوشبختانه، هیگینز و همکاران [۸۲] نشان داده‌اند که مدل‌های مولد عمیق به طور خاص قادر به یادگیری نمایش‌های جدا از هم هستند.

۳-۶- رمزگشایی با مدل‌های تولیدی

یک رویکرد پژوهشی امیدوارکننده، شامل ادغام روش‌های یادگیری عمیق در تحقیقات رمزگشایی مغز است. مدل‌های مولد عمیق مانند شبکه‌های خودرمزگذار متغیر [۱۸ ,۱۷] و شبکه-های مولد رقابتی [۱۹] در زمینه تولید تصاویر به موفقیت‌های بزرگی رسیده‌اند. اخیرا توجه روزافزونی متمرکز تحقیقات در زمینه بازسازی تصاویر بصری، با استفاده از مدل‌های عمیق تولیدی، شده است [۲۵–۲۰].

۳-۶-۱- روش های مبتنی بر شبکه‌های خودرمزگذار متغیر

شبکه‌های خودرمزگذار متغیر اولین بار در [۱۸ ,۱۷] ارائه شده‌اند. در این رویکرد مسئله تولید محتوا به یک مدل گرافیکی[۸] یا مدل گرافیکی احتمالاتی [۹](PGM) - مدلی احتمالاتی که توسط یک گراف، بیان‌گر ساختار وابستگی شرطی بین متغیرهای تصادفی است - تبدیل می‌شود. یک مدل

[۸] Graphical Model

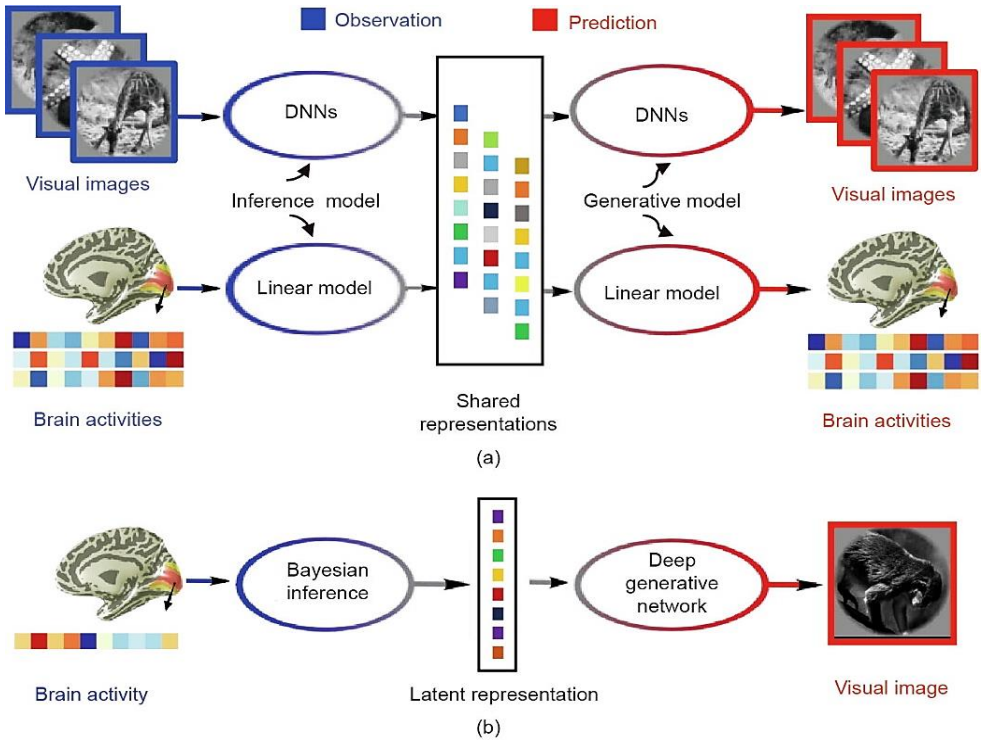
[۹] Probabilistic Graphical Model

خودرمزگذار متغیر، تشکیل شده از یک زیرشبکه رمزگذار پایین به بالا و یک زیرشبکه رمزگشایی بالا به پایین است. این دو زیرشبکه به صورت مشترک، به منظور ماکسیم کردن کرانه پایین لگاریتم درست‌نمایی، آموزش داده می‌شوند.

فعالیت‌های اخیر نشان داده‌اند که مدل‌های مبتنی‌بر شبکه‌های خودرمزگذار متغیر، قابلیت یادگیری بازنمایی‌های متمایزی را دارند که، با عوامل متفاوتی از تغییر در داده‌های ورودی مطابقت دارند [۸۱, ۸۳, ۸۴]. این موضوع در جهت رمزگذاری و رمزگشایی مغز بسیار مهم است، چرا که برخی از مفاهیم بصری آموخته شده توسط مدل‌های مبتنی بر شبکه‌های خودرمزگذار متغیر، توسط مغز انسان نیز درک می‌شوند. با الهام از این واقعیت، محققان استفاده از مدل‌های مبتنی بر شبکه‌های خودرمزگذار متغیر را در بازسازی تصویر از روی فعالیت‌های مغزی آغاز کرده‌اند [۲۱, ۲۲].

به عنوان مثال، Du و همکاران [۲۱] یک مدل چند نمایه مولد عمیق (DGMM)[۱۰] را برای بازسازی تصاویر درک شده از فعالیت‌های fMRI مغز پیشنهاد داده‌اند (شکل ۳-۳).

[۱۰] Deep Generative Multi-view Model



شکل ۳-۳: چارچوب مدل چند نمای مولد عمیق، به‌منظور رمزگشایی عصبی. (۱) آموزش مدل. (۲) بازسازی تصویر [۴۳].

DGMM را می‌توان به عنوان تعمیمی غیرخطی از تحلیل همبستگی متعارف بیزین خطی دانست. تحت چارچوب DGMM، روش‌های رمزگذاری و رمزگشایی به‌طور هم‌زمان توسط دو مدل تولیدی مجزا فرموله می‌شوند:

$$p_\theta\left(X \mid Z\right)=\prod_{i=1}^{N}\mathcal{N}\left\{x_i \mid \mu_x\left(z_i\right), diag\left[\sigma_x^2\left(z_i\right)\right]\right\} \qquad (۱-۲)$$

$$p\left(Y \mid Z\right)=\prod_{i=1}^{N}\mathcal{N}\left\{y_i \mid B^T z_i, \psi\right\} \qquad (۲-۲)$$

در روابط ذکر شده، $\mathcal{N}$ توزیع نرمال را نشان می‌دهد، $X \in \mathbb{R}^{D_x \times N}$ تصاویر بصری، $Y \in \mathbb{R}^{D_y \times N}$ فعالیت‌های fMRI برانگیخته شده، $p_\theta\left(X \mid Z\right)$ تابع درست‌نمایی تصاویر بصری با پارامتر شبکه عصبی θ، $p\left(Y \mid Z\right)$ تابع درست‌نمایی فعالیت‌های fMRI برانگیخته شده، ψ ماتریس کوواریانس، B وزن‌های متجسم فعالیت‌های fMRI برانگیخته شده، و $Z \in \mathbb{R}^{D_z \times N}$ نشان‌دهنده متغیرهای نهان مشترک بین تصاویر بصری و فعالیت‌های fMRI است. μ_x و σ_x^2 به ترتیب میانگین و کوواریانس

این توزیع نرمال را نشان می‌دهند که با تبدیلات غیرخطی مختلف با توجه به متغیرهای نهان بدست می‌آیند. مجموعه آموزشی متشکل از N نمونه جفت شده است، که می توان آن ها را با $(x_1, y_1),...,(x_n, y_n)$ نشان داد که در آن $x_i \in \mathbb{R}^{D_x}$ و $y_i \in \mathbb{R}^{Dy}$ برای $i = 1,..., N$.

به طور خاص، DGMM از یک فرآیند مولد مبتنی بر شبکه عصبی عمیق برای مدلسازی توزیع تصاویر بصری استفاده می‌کند، این در حالی است که، از یک فرآیند مولد خطی پراکنده برای مدلسازی توزیع داده‌های واکنش مغز استفاده می‌کند. از یک طرف، شبکه عصبی عمیقی که در اینجا به کار می‌رود می‌تواند به طور موثر ویژگی‌های سلسله مراتبی تصویر را ثبت کند که شبیه ساختار سیستم بصری غشائی مغز انسان هستند. از سوی دیگر، مدل مولد پراکنده خطی به‌کاررفته در اینجا نه تنها با اصل بیان پراکنده مغز انسان هم‌خوانی دارد، بلکه از بیش‌برازش داده‌های پاسخ مغز نیز جلوگیری می‌کند [۸۵]. توجه داشته باشید که این دو فرآیند مولد متغیرهای نهان یکسانی دارند. بنابراین، در مرحله تست، استفاده از این فرآیندها امکان استنتاج تصویر از واکنش مغز از طریق متغیرهای نهان مشترک را ممکن می‌سازد.

در حقیقت چارچوب DGMM می‌تواند روابط نگاشت دوسویه بین تصاویر بصری و فعالیت‌های fMRI متناظر را ثبت کند. به لطف معماری بیزین متغیر خودرمزگذار DGMM را می توان به‌طور موثر با استفاده از استنتاج متغیر میدان متوسط که مشابه راه‌حل شبکه‌های خودرمزگذار متغیر کلاسیک است، بهینه سازی کرد.

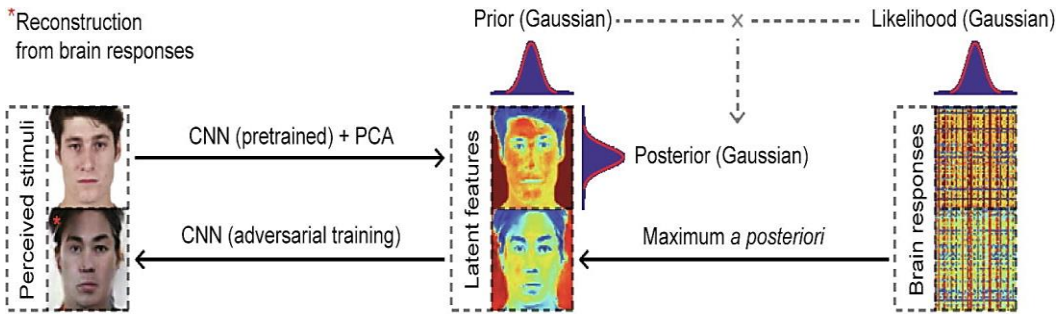
۳-۶-۲- روش های مبتنی بر شبکه‌های مولد رقابتی

شبکه‌های مولد رقابتی برای اولین بار در [۱۹] پیشنهاد شده‌اند. شبکه مولد رقابتی ساده، یک مدل نظارت نشده است که، از یک بردار نویز تعدادی تصویر تولید می‌کند. ایده آموزش رقابتی، از تئوری بازی نشأت می‌گیرد. به‌این صورت‌که، دو شرکت‌کننده با هم رقابت می‌کنند تا هر دو پیشرفت کنند. پیکربندی معمول در یک شبکه مولد رقابتی، شامل یک مولد و یک تفکیک‌کننده

می‌شود. وظیفه مولد، ساخت تصاویر از روی نویز است. به‌شکلی که تفکیک‌کننده را مجاب کند که این تصویر مربوط به یک صحنه از دنیای واقعی است. در همین حال، تفکیک‌کننده تلاش می‌کند تا بین داده‌های تولید شده و داده‌های واقعی تمییز بدهد. هنگامی‌که تعادل نَش[۱۱] برقرار شود، مولد، توزیع تصاویر دنیای واقعی را یاد می‌گیرد شبکه‌های مولد رقابتی در کاربردهای متنوعی مانند تولید تصاویر [۸۶]، انتقال تصویر به تصویر [۸۷] و ترکیب متن با تصویر [۸۸, ۸۹] به‌کار می‌روند.

برخلاف یک شبکه خودرمزگذار متغیر، شبکه مولد رقابتی یک مدل مستقل از درست‌نمایی است. به‌این صورت که هیچ فرض قبلی‌ای راجع‌به توزیع داده‌ها صورت نمی‌گیرد و توزیع داده‌ها تماما در طول آموزش رقابتی یاد گرفته می‌شود. این یک مشخصه مطلوب در رمزگذاری و رمزگشایی عصبی است. شبکه مولد رقابتی، معمولا نیازمند یک جریان دقیقی از اطلاعات معنایی، هم در مولد و هم در تفکیک‌کننده است. البته اطلاعات معنایی و مفید در سیگنال وابسته به سطح اکسیژن خون با انبوهی از نویزها ترکیب شده است، که این، چالش بزرگی برای آموزش مدل است.

پژوهش اخیر در زمینه رمزگشایی مغز [۲۰]، ترکیبی از استنباط احتمالی[۱۲] با آموزش رقابتی ارائه کرده است، که در بازسازی چهره‌های درک شده، از روی فعالیت‌های مغزی به کار برده می‌شوند (شکل ۳-۴).



شکل ۳-۴: رمزگشایی عصبی رقابتی عمیق. ترکیب استنباط احتمالی با یادگیری رقابتی، این روش می تواند به وضوح تصویر مربوطه از چهره را از روی فعالیت مغز بازسازی کند [۲۰].

[۱۱] Nash equilibrium

[۱۲] Probabilistic inference

اگر فرض کنیم $x \in \mathbb{R}^{h\times w\times c}$ تصویر دیده شده، $z \in \mathbb{R}^{p}$ ویژگی‌های نهان، $y \in \mathbb{R}^{q}$ پاسخ متناظر از طرف مغز و $\varphi \in \mathbb{R}^{h\times w\times c} \to \mathbb{R}^{p}$ یک مدل از ویژگی نهانی چنان‌که داشته باشیم $z = \varphi(x)$ و $x = \varphi^{-1}(z)$، باشند. سپس تصاویر دیده شده و درک شده را می‌توان از روی پاسخ‌های مغز و به کمک رابطه زیر بازسازی کرد:

$$\hat{x} = \varphi^{-1}\left[\arg\max_{z} p(z \mid y)\right] \qquad (۳-۲)$$

که در آن $p(z \mid y)$، احتمال پسین متغیرهای نهان است. رابطه ۲-۳ را می‌توان با استفاده از قضیه بیز بازنویسی کرد:

$$\hat{x} = \varphi^{-1}\left\{\arg\max_{z}\left[p(y \mid z)\, p(z)\right]\right\} \qquad (۴-۲)$$

که در آن $p(y \mid z)$، تابع درست‌نمایی و $p(z)$، توزیع پیشین متغیرهای نهان است. محققان در ابتدا به‌طور شهودی ویژگی‌های نهان را از روی پاسخ‌های مشاهده شده از مغز، به‌وسیله یک تخمین پسین بیشینه، رمزگشایی کردند. سپس با به‌کارگیری یادگیری رقابتی، تصاویر درک شده را، با توجه به ویژگی‌های نهان رمزگشایی شده، تولید می‌کنند.

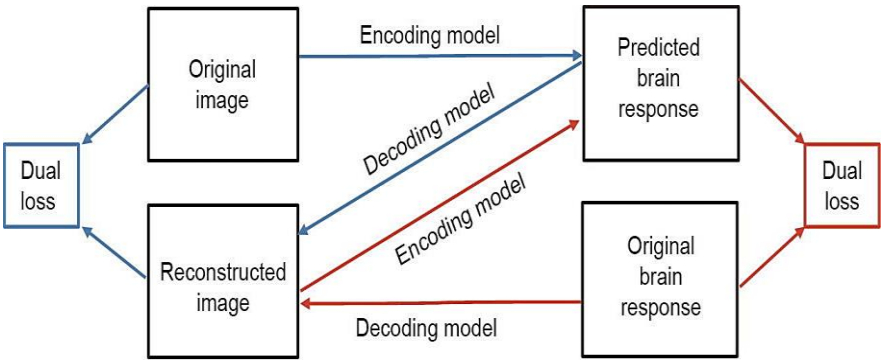
این رویکرد دومرحله‌ای رمزگشایی مغز قادر است به‌طور دقیقی بازسازی چهره‌های درک شده را از روی پاسخ‌های مغزی، تولید کند. در تحقیقات جدیدتر، محققان تلاش کردند تا با استفاده از سیگنال‌های fMRI، تصاویر طبیعی را بازسازی کنند [۲۵–۲۳]. آن‌ها این کار را به‌وسیله شبکه‌ها مولد رقابتی‌ای که با دیتاست‌های بزرگ‌مقیاس تصاویر، پیش آموزش داده شده بودند، انجام دادند.

۳-۷- بهبود رمزگذاری و رمزگشایی مغز با یادگیری دوگانه

روش‌های رمزگذاری و رمزگشایی مغز مبتنی‌بر داده، به‌منظور آموزش مدل خاص یک سوژه مجزا، اغلب مستلزم جمع‌آوری تعداد زیادی نمونه از داده‌های جفت شده (تحریک – پاسخ) هستند. با این حال، در بسیاری از مطالعات رمزگذاری و رمزگشایی، جمع‌آوری چندهزار نمونه داده همراه با

نویز حداکثر از یک سوژه امکان پذیر است. بنابراین برای بهبود توانایی تعمیم مدل‌های رمزگذاری و رمزگشایی، لازم است که از نمونه داده‌های جفت نشده در مقیاس بزرگ (به‌عنوان مثال، تصاویر بصری) استفاده خوبی داشته باشیم.

با الهام از یادگیری دوگانه پیشنهادی برای ترجمه ماشینی[۱۳] [۹۰, ۹۱]، پیشنهاد می‌کنیم که این امکان وجود دارد که مدل‌های رمزگذاری و رمزگشایی را به‌طور همزمان با به حداقل رساندن خطای بازسازی ناشی از مدل نگاشت دوگانه آموزش دهیم. در شکل ۳-۵ مدل‌های رمزگذاری و رمزگشایی، یک حلقه بسته را شکل می‌دهند، و شرایط به‌کارگیری یادگیری دوگانه را فراهم می‌کنند.



شکل ۳-۵: بهبود رمزگذاری و رمزگشایی مغز با یادگیری دوگانه [۴۳].

به‌طور خاص، خطای بازسازی محاسبه شده روی داده‌های جفت نشده (به عنوان مثال، تصاویر) بازخورد مفیدی را برای آموزش مدل نگاشت دوگانه ایجاد می‌کند. تحت این چارچوب یادگیری دوگانه، امکان استفاده از تصاویر جفت نشده مقیاس بزرگ برای بهبود توانایی تعمیم مدل‌های رمزگذاری و رمزگشایی وجود دارد.

در حقیقت، یادگیری دوگانه یک چارچوب کلی برای یادگیری نگاشت های دوگانه از یک حوزه داده M_d به حوزه داده دیگر N_d را فراهم می‌سازد. برای $M_d \to N_d$، هدف اصلی یادگیری یک نگاشت رمزگشای E است؛ بگونه ای که با استفاده از خطای رقابتی، توزیع $E(M_d)$ از توزیع

[۱۳] Machine translation

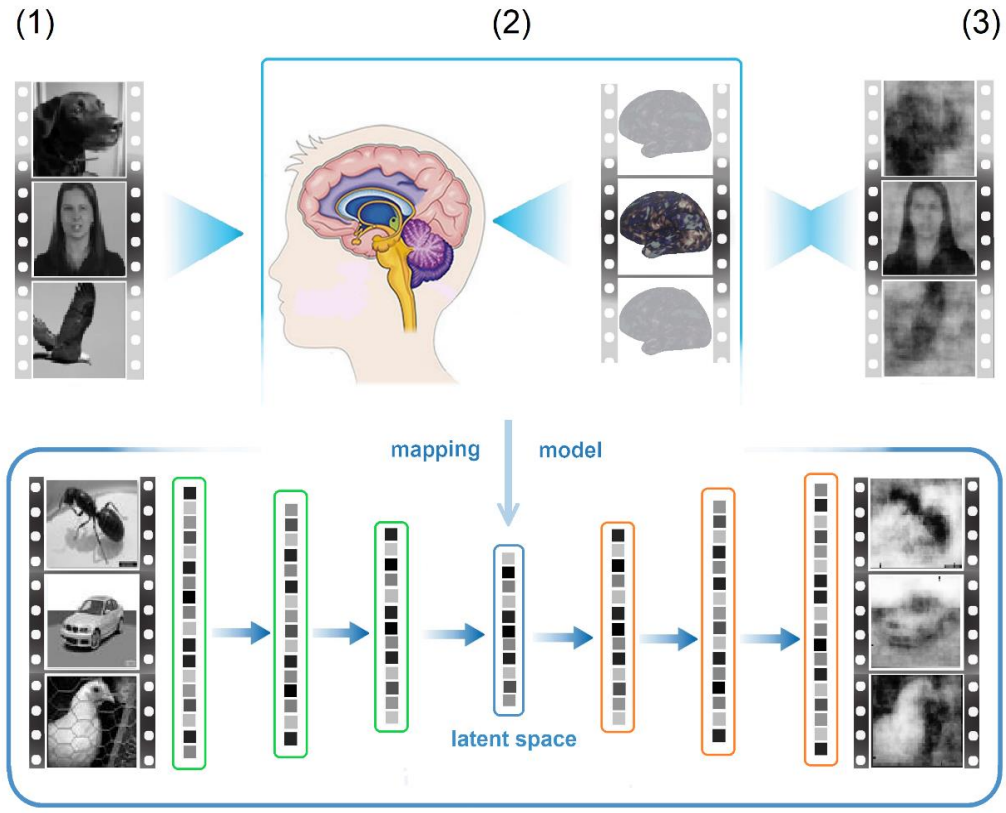
N_d قابل تشخیص نباشد. به طور مشابه، برای $N_d \to M_d$ هدف، یادگیری یک نگاشت رمزگشای D است به گونه‌ای که توزیع $D(N_d)$ با استفاده از یک خطای رقابتی دیگر از توزیع M_d قابل‌تشخیص نباشد. به‌طور خاص، برای داده‌های جفت شده، امکان ترکیب این دو خطای رقابتی و خطای پایداری چرخه (خطای دوگانه) برای تحقق $D\left[E(M_d)\right] \approx M_d$ و $E\left[D(N_d)\right] \approx N_d$ وجود دارد.

۴- فصل چهارم

روش پیشنهادی پژوهش

۴-۱- مقدمه

در این فصل به معرفی روش پیشنهادی این پژوهش به منظور بازسازی تصاویر درک شده توسط کاربر از روی داده‌های fMRI می‌پردازیم. بر این اساس مدلی، بر مبنای شبکه‌های عصبی ارائه می‌شود که قادر به رمزگشایی مغز باشد، به‌گونه‌ای که تصاویر قابل درکی از آن‌چه که انسان مشاهده کرده است، بازگرداند. همان‌طور که در شکل ۴-۱ نشان داده شده است.



شکل ۴-۱: بازسازی تصاویر درک شده توسط کاربر از روی داده‌های رزونانس مغناطیسی عملکردی. (۱) درک تصاویر مشاهده شده توسط کاربر و ایجاد بازنمایی‌هایی در ذهن سوژه. (۲) ثبت تصاویر رزونانس مغناطیسی عملکردی از سوژه در حین مشاهده تصاویر و مدلسازی عملکرد رمزگشایی مغز براساس داده‌های جمع‌آوری شده. (۳) بازسازی تصاویر درک شده مطابق مدل ارائه شده.

جهت محقق ساختن چنین مدلی، در ابتدا یک شبکه عصبی AAE را با بیش از ۱۱ هزار تصویر طبیعی رویت نشده توسط سوژه‌ها، آموزش می‌دهیم. فضای‌نهان این ساختار، توصیفی معنادار (۱۰۰ بُعدی) از هر تصویر را ارائه می‌دهد. سپس نگاشتی از داده‌های fMRI یکی از سه سوژه انسانی به فضای‌نهان ۱۰۰ بُعدی شکل می‌دهیم، که بر اساس آن داده‌های fMRI به کدهای نهان ۱۰۰

بُعدی تبدیل می‌شوند. در گام تست، از نگاشت شکل گرفته استفاده می‌کنیم و الگوهای fMRI هر سه سوژه انسانی را به کدهای نهان تبدیل می‌کنیم، در نهایت کدها با استفاده از ساختار رمزگشای شبکه AAE، به بازنمایی‌های قابل درکی از تصاویر تست، بدل می‌شوند. مدل رمزگشایی پیشنهادی، با استفاده از داده‌های مجزا، آموزش می‌بیند و تست می‌شود.

رویکرد پیشنهادی در این پژوهش با درنظر گرفتن تصاویر در سطح خاکستری معرفی می‌گردد. اعمال چنین فرضی با هدف کاهش فرآیندهای پردازشی و ساده ساختن ساختار شبکه عصبی AAE صورت می‌پذیرد.

در ادامه مجموعه داده به‌کار گرفته شده و همچنین جزییات مدل پیشنهادی شرح داده می‌شود. رویکرد پیشنهادی در محیط نرم‌افزار MATLAB پیاده‌سازی شده است.

۴-۲- پایگاه داده

مجموعه داده مورد استفاده در این پژوهش شامل ۱۱/۵ ساعت از داده‌های fMRI، جمع‌آوری شده از ۳ سوژه انسانی است. در حین اخذ داده‌ها سوژه‌ها در حال تماشای تقریبا ۹۷۲ کلیپ ویدئویی مختلف، شامل اشیاء، صحنه‌ها و کنش‌های متنوع بوده‌اند. این مجموعه داده در آزمایشگاه IBI دانشگاه Purdue ثبت شده است و نسبت به سایر نمونه‌های موجود در سایر مطالعات، اندازه و پوشش گسترده تری دارد و به صورت دسترسی آزاد در سایت دانشگاه Purdue (https://engineering.purdue.edu/libi/lab/Resource.html) قرار داده شده است [۹۲].

۴-۲-۱- سوژه‌ها و آزمایش‌ها

سه داوطلب سالم (زن، سن ۲۲-۲۵) با بینایی طبیعی در این مطالعه شرکت کردند. به هر نفر آموزش داده شد که ضمن تمرکز بر یک ناحیه ثابت ($^{\circ}$۰/۸ × $^{\circ}$۰/۸) یک سری کلیپ ویدئویی بی‌صدا شامل تصاویر طبیعی ($^{\circ}$۲۰/۳ × $^{\circ}$۲۰/۳) را تماشا کنند.

۳۷۴ کلیپ ویدئویی به صورت پیوسته (با نرخ‌فریم ۳۰ فریم در ثانیه) در قالب یک فیلم آموزشی ۲/۴ ساعته، به طور تصادفی به ۱۸ بخش ۸ دقیقه‌ای تقسیم شده است. ۵۹۸ کلیپ ویدئویی مختلف نیز در قالب یک فیلم تست ۴۰ دقیقه‌ای گنجانده شده است و به طور تصادفی به ۵ بخش ۸ دقیقه‌ای تقسیم شده است.

کلیپ‌های ویدئویی در فیلم آموزش با کلیپ‌های ویدئویی در فیلم تست با هم متفاوت هستند. کلیپ‌های ویدئویی از سایت VideoBlocks و YouTube انتخاب شده‌اند تا متنوع و در عین حال نماینده‌ای از تجربیات بصری در زندگی واقعی باشند. به عنوان مثال، کلیپ‌های ویدئویی افرادِ در حال فعالیت، حیواناتِ در حال حرکت، چشم‌اندازهای طبیعی، صحنه‌های سرباز یا سرپوشیده و غیره را نمایش می‌دهند. هر سوژه، دو مرتبه ویدئویِ آموزشی و ده مرتبه ویدئویِ تست را در روزهای مختلفِ آزمایش تماشا کرده است. هر آزمایش شامل چندین نشست به مدت ۸ دقیقه و ۲۴ ثانیه بوده است. در طول هر نشست، یک بخش ویدئوِ ۸ دقیقه‌ای ارائه شده است. اولین و آخرین فریم، هر کدام به مدت ۱۲ ثانیه به صورت یک تصویر ایستا نمایش داده می‌شده است. ترتیب بخش‌های ویدئوها به صورت تصادفی و غیر متعادل بوده است. محرک‌های بصری با استفاده از جعبه‌ابزار Psychophysics3، روی صفحه نمایش با وضوح ۸۰۰ × ۶۰۰ به سوژه‌ها نمایش داده شده اند.

۴-۲-۲- جمع‌آوری داده‌ها

داده‌های MRI و fMRI، T_1[۱] و T_2[۲] وزن‌دار در یک سیستم MRI ۳ تسلا جمع‌آوری شده‌اند. داده‌های fMRI با وضوح مکانی همگن[۳] ۳/۵ میلی‌متری و وضوح زمانی[۴] ۲ ثانیه با استفاده از روش

[۱] Spin-lattice relaxation time

[۲] Spin-spin relaxation time

[۳] Ispatial resolution

[۴] Temporal resolution

تصویربرداری سطح بازتابی (EPI)[۱] تک‌برش (SS)[۲]، جمع‌آوری شده‌اند (۳۸ برش محوری[۳] با ضخامت[۴] ۳/۵ میلی‌متر و وضوح درون صفحه‌ای[۵] ۳/۵×۳/۵ میلی‌متر مربع، زمان تکرار (TR)[۶] ۲۰۰۰ میلی ثانیه، زمان اکو (TE)[۷] ۳۵ میلی ثانیه، زاویه تلنگر[۸] ۷۸° و میدان دید[۹] ۲۲ × ۲۲ سانتی‌متر مربع).

هنگام آموزش و تست مدل رمزگشایی پیشنهادی، سیگنال‌های fMRI قشری[۱۰]، مقدار متوسط بین چند تکرار هستند: ۲ تکرار برای فیلم آموزشی و ۱۰ تکرار برای فیلم تست. ده تکرار فیلم تست نیز، به ما امکان می‌دهد پاسخ‌های مغز را با نسبت‌های سیگنال به نویز (SNR)[۱۱] بالاتری بدست آوریم، چرا که فعالیت‌های خودبه‌خودی و بی‌اختیار یا نویزهای غیرمرتبط با محرک‌های بینایی به طور موثر با میانگین‌گیری این تعداد تکرار حذف می‌شوند. اگرچه تکرارهای بیشتر برای فیلم آموزشی هم می‌توانست به افزایش SNR داده‌های آموزشی نیز کمک کند، اما به دلیل طولانی بودن زمان فیلم آموزشی به تعداد دفعات تکرار فیلم تست، تکرار نشدند.

۴-۲-۳ پیش‌پردازش داده‌ها

مجموعه داده معرفی شده با استفاده از نرم‌افزار Connectome Workbench پیش‌پردازش شده و به صورت آزاد در اختیار بود. اما به‌منظور حفظ یگانگی پژوهش، مجموعه تصاویر fMRI مجددا به منظور استخراج سری‌های زمانی مربوط به ناحیه بینایی با استفاده از نرم‌افزار SPM12

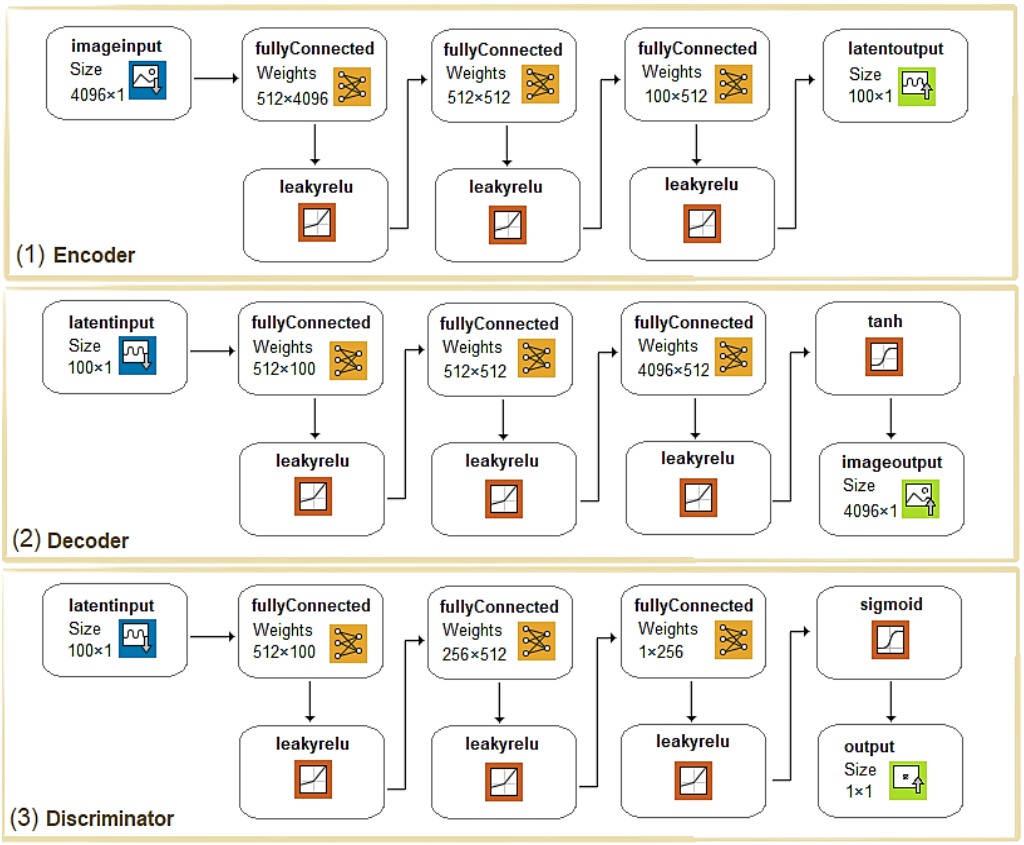
---

[۱] Echo Planar Imaging
[۲] Single Shot
[۳] Axial slices
[۴] Slice thickness
[۵] In-plane resolution
[۶] Repetition Time
[۷] Echo Time
[۸] Flip angle
[۹] Field of view
[۱۰] Cortical
[۱۱] Signal to Noise Ratios

پیش‌پردازش می‌شوند و ادامه روند پیشنهادی بر روی سیگنال‌های fMRI ناحیه بینایی صورت می‌گیرد.

۴-۳- ساختار شبکه خودرمزگذار رقابتی پیشنهادی

از آنجا که هدف اصلی پژوهش، با دست‌یابی به کدهای نهان تصاویر محقق می‌شود، و اصلاح وزن در شبکه‌های خودرمزگذار رقابتی نیز براساس تفکیک کدهای نهان صورت می‌پذیرد، این نوع شبکه در این پژوهش مورد استفاده قرار گرفته است. شبکه خودرمزگذار رقابتی پیشنهادی، شامل سه زیرشبکه معرفی شده در بخش ۲-۷ است. زیرشبکه‌ها با ساختار نرونی و با اتصال کامل درنظر گرفته شده‌اند. در شکل ۴-۲ جزئیات ساختار هر زیرشبکه قابل مشاهده است. لازم به ذکر است، تعداد لایه‌ها و ابعاد هر لایه در زیرشبکه‌ها، از جمله بُعد فضای نهان در این ساختار، به‌صورت تجربی و با توجه به دست‌یابی به عملکرد قابل قبول، انتخاب شده‌اند.



شکل ۴-۲: معماری شبکه خودرمزگذار رقابتی به‌کار گرفته شده. (۱) معماری زیرشبکه رمزگذار. (۲) معماری زیرشبکه رمزگشا. (۳) معماری زیرشبکه تفکیک‌کننده

در این ساختار ماتریس تصویر ورودی با ابعاد ۶۴ × ۶۴ پیکسل، به صورت برداری (۴۰۹۶ بُعدی) تغییر شکل داده می‌شود و در اختیار زیرشبکه رمزگذار (۴ لایه‌ای) قرار می‌گیرد. رمزگذار براساس مدل گاوسی پسین، به ازای هر تصویر ورودی یک کد نهان ۱۰۰ بُعدی بازمی‌گرداند.

کد نهان ایجاد شده همراه با یک نمونه کد ۱۰۰ بُعدی تولید شده بر اساس توزیع پیشین (توزیع گاوسی) به عنوان ورودی به زیرشبکه تفکیک‌کننده (۴ لایه‌ای) داده می‌شود. در مرحله تنظیم از آموزش شبکه، زیرشبکه تفکیک‌کننده خود را با استفاده از روش گرادیان کاهشی تصادفی، به‌روزرسانی می‌کند تا کدهای نهان تولید شده بر اساس توزیع گاوسی را از نمونه‌های تولید شده توسط رمزگذار جدا کند. اگر $\hat{P}_{fake}$ را احتمال خروجی تفکیک‌کننده به ازای کد تولید شده توسط رمزگذار درنظر بگیریم، و $\hat{P}_{real}$ احتمال خروجی تفکیک‌کننده به ازای کد تولید شده بر اساس توزیع گاوسی باشد، مقدار خطا در زیرشبکه تفکیک‌کننده به شرح زیر تعریف می‌شود:

$$loss_{\mathrm{D}} = -\frac{1}{2N}\sum_{i=1}^{N}\log\left(\hat{P}_{real_i}+\varepsilon\right)+\frac{1}{N}\sum_{i=1}^{N}\log\left(1-\hat{P}_{fake_i}+\varepsilon\right) \qquad (۱-۴)$$

از طرفی، کد نهان ایجاد شده توسط رمزگذار به عنوان ورودی به زیرشبکه رمزگشا داده می‌شود، تا در خروجی به برداری با ابعاد ۱ × ۴۰۹۶ دست یابیم. با تبدیل ابعاد بردار به دست آمده، به ماتریس تصویر بازسازی شده هم بُعد تصویر ورودی می‌رسیم.

به‌روزرسانی زیرشبکه‌های رمزگذار و رمزگشا نیز براساس روش گرادیان کاهشی تصادفی انجام می‌شود، به‌گونه‌ای که هم زیرشبکه رمزگذار بتواند تفکیک‌کننده را به اشتباه بیندازد و هم اینکه زیرشبکه‌های رمزگذار و رمزگشا خطای بازسازی ورودی‌ها، یا همان مجموع مربعات خطا را به حداقل برسانند. اگر درنظر بگیریم که x و $\hat{x}$ به ترتیب تصاویر ورودی رمزگذار و خروجی رمزگشا باشند، λ_1 و λ_2 نیز پارامتر تاثیر هر ترم خطا باشند، مقدار خطا در این دو زیرشبکه به شرح زیر تعریف می‌شود:

$$loss_A = \lambda_1 \frac{1}{2N} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(x_{ij} - \hat{x}_{ij} \right)^2 - \lambda_2 \frac{1}{N} \sum_{i=1}^{N} \log\left(1 - \hat{P}_{fake_i} + \varepsilon\right) \quad (۲-۴)$$

در واقع پس از اتمام مراحل آموزش، زیرشبکه رمزگشا، یک مدل مولدی را تعریف می‌کند که قادر است توزیع پیشین تحمیل شده، را به توزیع داده‌ها نگاشت کند. و زیرشبکه رمزگذار نیز مدلی مولد را ارائه می‌دهد که توزیع داده‌ها را به توزیع پیشین تبدیل می‌کند.

به منظور آموزش شبکه خودرمزگذار رقابتی از ۱۱۲۵۰ تصویر طبیعی برچسب گذاری شده توسط انسان، شامل ۱۵ گروه کلیِ موجود در فیلم‌های آموزش و تست استفاده می‌شود. ۱۵ گروه از تصاویر طبیعی، تحت عناوین؛ محیط سرپوشیده، محیط سرباز، انسان، چهره، پرنده، حشره، حیوانات آب‌زی، حیوانات خشکی‌زی، گل، میوه، چشم‌اندازهای طبیعی، ماشین، هواپیما، کشتی و ورزش کردن، برچسب‌گذاری شده‌اند. این گروه‌بندی‌ها محتوای مشترکی را در هر دو فیلم‌های آموزش و تست، پوشش می‌دهند. در ادامه با استفاده از مجموعه‌ای متفاوت، شامل ۱۵۰۰ تصویر دارای برچسب، تست شبکه انجام می‌شود. تصاویر مجموعه آموزش و تست همه به صورت تصادفی و به طور مساوی از ۱۵ گروه فوق در دیتابیس‌های ImageNet و COCO نمونه برداری شدند و پس از آن نیز، بررسی بصری به منظور جایگزینی تصاویر دارای برچسب اشتباه انجام شد.

آموزش شبکه در ۲۵۰ دوره[۱] و در هر دوره به ازای نمونه‌های ۲۰تایی از تصاویر انجام شد. همان‌طور که بیان شد در طی فرآیند آموزش شبکه خودرمزگذار رقابتی سه زیرشبکه آموزش داده می‌شوند: رمزگذار؛ آموزش می‌بیند، تصاویر طبیعی ورودی را به کدهای فضای‌نهان نگاشت کند، به گونه‌ای که هر کد از فضای‌نهان توصیفی معنادار از هر تصویر را ارائه دهد. رمزگشا؛ می‌آموزد که کدهای نهان ۱۰۰ بُعدیِ فضای‌نهان را به تصاویر طبیعی قابل قبولی تبدیل کند. تفکیک‌کننده، که فقط در مرحله آموزش استفاده می‌شود، یاد می‌گیرد که نمونه‌های واقعی (تولید شده بر اساس توزیع پیشین) را از نمونه های تولید شده (کدهای نهان محاسبه شده توسط خودرمزگذار) جدا کند.

[۱] Epoch

پس از اتمام آموزشِ شبکه خودرمزگذار رقابتی، زیرشبکه تفکیک‌کننده کنار گذاشته می‌شود و تست شبکه با ۱۵۰۰ تصویر دارای برچسب صورت می‌گیرد.

در ادامه و به‌طور خاص، از زیرشبکه رمزگذار این ساختار برای تولید کدهای نهان ۱۰۰ بُعدی به‌ازای هر فریم (در مجموع ۴۳۲۱ فریم) از مجموعه فیلم آموزشی نشان داده شده به یکی از سوژه‌های انسانی، حین اخذ داده‌های fMRI، استفاده می‌کنیم. سپس این کدها به عنوان ماتریس طراحی، ذخیره می‌شوند و به منظور آموزش نگاشتی مناسب، جهت تبدیل سیگنال‌های fMRI یکی از سه سوژه انسانی به فضای‌نهان ۱۰۰ بُعدی، مورد استفاده قرار می‌گیرند. پس از دست‌یابی به کدهای نهان نگاشت شده از روی سیگنال‌های fMRI، زیرشبکه رمزگشا آموزش داده شده را مورد استفاده قرار می‌دهیم. به‌گونه‌ای که این‌بار نگاشتی از فضای نهان به تصاویر بازسازی شده خواهیم داشت.

به‌منظور ارزیابی کمی عملکرد شبکه، شباهت تصاویر تستِ ورودی و تصاویر بازسازی شدهِ خروجی، سنجیده می‌شود. این شباهت براساس دو معیار متکی‌بر مرجع محاسبه می‌شود.

اولین معیار مورد استفاده، معیار شباهت ساختاری چندمقیاسه[۱] (Ms-SSIM) است. بدیهی است که، مقادیر پیکسل‌ها بسته به موقعیتشان در تصویر، تاثیر متفاوتی بر روی چشم انسان ایجاد می‌کنند. از جمله معیارهای جدیدی که در آن‌ها مشخصه‌های ساختاری تصویر بر اساس اطلاعات آماری تجربی به دست می‌آید، می‌توان به خانواده معیارهای شباهت ساختاری اشاره کرد. شباهت ساختاری چندمقیاسه فُرم پیشرفته‌تری از شباهت ساختاری[۲] و یا به عبارتی همان شباهت ساختاری تک‌مقیاسه (SSIM) است. برای دو تصویر y و x، قیاس از نظر درخشندگی[۳]، کنتراست[۴] و ساختار به‌ترتیب به‌صورت زیر قابل محاسبه هستند:

[۱] Multi-scale Structural Similarity
[۲] Structural Similarity
[۳] Luminance
[۴] Contrast

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \qquad (۳-۴)$$

$$c(x,y)=\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \qquad (۵-۴)$$

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \qquad (۶-۴)$$

در رابطه‌های ذکر شده، μ_x و μ_y به‌ترتیب مقادیر میانگین تصاویر x و y هستند. σ_x^2 و σ_y^2 نیز به‌ترتیب مقادیر واریانس تصاویر x و y تعریف می‌شوند. σ_{xy} کواریانس محاسبه شده بین این دو تصویر است. C_1، C_2 و C_3 پارامترهای تنظیم رابطه هستند و به شرح زیر تعریف می‌شوند:

$$\begin{aligned} C_1&=(k_1L)^2\\ C_2&=(k_2L)^2\\ C_3&=C_2/2 \end{aligned} \qquad (۷-۴)$$

L در رابطه معادل با محدوده تغییرات مقادیر پیکسل‌ها است. ضرایب k_1 و k_2 نیز به ترتیب برابر با ۰/۰۱ و ۰/۰۳ در نظر گرفته می‌شوند. در نهایت شباهت ساختاری به‌صورت زیر تعریف می‌شود:

$$SSIM(x,y)=\left[l(x,y)\right]^{\alpha}\left[c(x,y)\right]^{\beta}\left[s(x,y)\right]^{\gamma} \qquad (۸-۴)$$

روش چندمقیاسه یک روش مناسب برای ترکیب جزئیات تصویر در رزولوشن‌های مختلف است. در این روش رابطه ۴-۸ به شکل زیر بازنویسی می‌شود:

$$Ms-SSIM(x,y)=\left[l_M(x,y)\right]^{\alpha_M}\prod_{j=1}^{M}\left[c_j(x,y)\right]^{\beta_j}\left[s_j(x,y)\right]^{\gamma_j} \qquad (۹-۴)$$

در این روش، تصویر اصلی به عنوان مقیاس ۱ و بالاترین مقیاس، با مقیاس M فهرست می‌شوند. در مقیاس jام، مقایسه کنتراست و مقایسه ساختار به ترتیب با $c_j(x,y)$ و $s_j(x,y)$ محاسبه می‌شوند. مقایسه درخشندگی تنها در مقیاس M محاسبه می‌شود.

دومین معیار مورد استفاده، معیار ضریب همبستگی[۱] (CC) است. معیاری عددی، که به نوعی بیان‌گر رابطه آماری بین دو تصویر است. این معیار به شرح زیر بین دو تصویر x و y قابل محاسبه است:

$$CC(x,y)=\frac{\sum_m\sum_n\left(x_{mn}-\mu_x\right)\left(y_{mn}-\mu_y\right)}{\sqrt{\left(\sum_m\sum_n\left(x_{mn}-\mu_y\right)^2\right)\left(\sum_m\sum_n\left(y_{mn}-\mu_y\right)^2\right)}} \qquad (۱۰-۴)$$

در اینجا، μ_x و μ_y به‌ترتیب مقادیر میانگین تصاویر x و y هستند.

هر دو معیار درنظر گرفته شده به‌ازای دو تصویر کاملا یکسان، مقداری برابر با ۱ خواهند داشت. از رو هرچه مقدار به ۱ نزدیک‌تر باشد، بازسازی بهتری داشته انجام گرفته است.

۴-۴- نگاشت سیگنال‌های fMRI

در گام سوم، پس از انجام دو مرحله پیش‌پردازش داده‌های fMRI و آموزش AAE، نگاشتی طراحی می‌شود تا سیگنال‌های fMRI را به کدهای نهان تبدیل کند. کدهای نهانی که به عنوان ورودی به زیرشبکه رمزگشا آموزش دیده شده در قبل، داده می‌شوند و بر این اساس در خروجی بازنمایی قابل درکی از تصاویر طبیعی تحریک را خواهیم داشت. این نگاشت با استفاده از دو رویکرد رگرسیون خطی چندمتغیره[۲] و شبکه عصبی، برآورد می‌شود و آموزش آن تنها با استفاده از داده‌های fMRI مربوط به یکی از سه سوژه انسانی انجام صورت می‌پذیرد. این امر سبب می‌شود تا مدل رمزگشایی نهایی، تعمیم‌پذیری بین انسانی بالاتری داشته باشد.

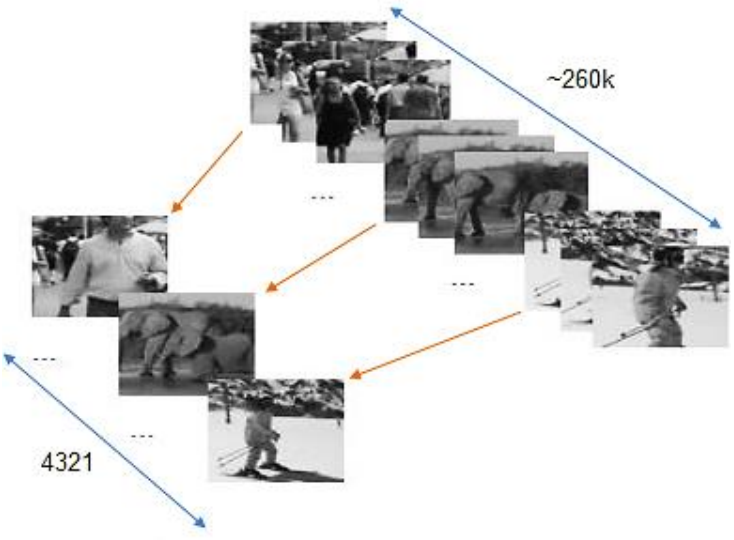
---

[۱] Correlation Coefficient

[۲] Multivariate linear regression

جهت آموزش چنین نگاشتی نیازمند کدهای نهان مربوط به تصاویر تحریک هستیم. بدین منظور لازم است، ابتدا ویدئوهای تحریک، فریم‌بندی شوند و سپس با استفاده از زیرشبکه رمزگذار آموزش دیده شده در قبل، به کدهای نهان مدنظر برسیم.

هنگام استخراج فریم‌های فیلم آموزشی ۲/۴ ساعته، ابعاد تصاویر برابر با ابعاد تصاویر ورودی در شبکه خودرمزگذار رقابتی (۶۴ پیکسل × ۶۴ پیکسل) درنظر گرفته می‌شوند. از آنجا که نرخ‌فریم فیلم آموزشی برابر ۳۰ فریم در ثانیه است، هر بخش ۸ دقیقه‌ای شامل ۱۴۴۰۱ تصویر می‌شود و به صورت کلی، به‌ازای همه ۱۸ بخش، نزدیک به ۲۶۰ هزار تصویر خواهیم داشت. به منظور مطابقت تعداد فریم‌ها با تعداد حجم‌ها[1] در داده‌های fMRI، فریم‌ها با فرکانس ۳۰×۲ نمونه‌برداری می‌شوند، که در آن ۲، برابر با نرخ نمونه‌برداری هنگام ثبت داده‌ها است. نهایتا به ازای ۲/۴ ساعت فیلم آموزشی، ۴۳۲۱ فریم، تصویر طبیعی خواهیم داشت (شکل ۴-۳).



شکل ۴-۳: فریم‌بندی ویدئو تحریک آمورشی.

[1] Volumes

۴-۴-۱- مدل رگرسیون خطی چندمتغیره

می‌توان گفت، مدل رگرسیون تعریف شده معیاری است در جهت انتخاب وکسل‌های برانگیخته شده در اثر مشاهده تصاویر تحریک. زمان‌های اخذ fMRI و یا به عبارتی حجم‌ها، نمونه‌های ورودی و خروجی مدل را تعیین می‌کنند. مدل رگرسیون خطی چندمتغیره به شرح زیر توصیف می‌شود:

$$Y = XW + \varepsilon \qquad (۱۱-۴)$$

در اینجا، X بیان‌گر ماتریس طراحی یا همان کدهای نهان مربوط به فریم‌های ویدئو تحریک است. ماتریسی $N \times (l+1)$ بُعدی، که در آن N تعداد حجم‌ها و l طول هر بردار نهان است. آخرین ستون این ماتریس، برداری است که مقادیر تمام عناصر آن برابر یک و به منظور اعمال مقدار ثابت در مدل رگرسیون، در نظر گرفته شده است.

ماتریس Y، متغیر پاسخ نیز، مربوط به سیگنال‌های fMRI است. این ماتریس ابعادی برابر $N \times v$ دارد که در آن v، بیان‌گر تعداد وکسل‌ها و N تعداد حجم‌ها است.

W، این ماتریس دارای $l+1$ سطر و v ستون است که سطر W_0 مقدار ثابت در مدل را نشان می‌دهد. این پارامترها در صورت برآورد، همان ضرایب مدل رگرسیونی[۱] هستند. مقدار این ضرایب، حساسیت متغیر پاسخ را به هر یک از متغیرهای مستقل نشان می‌دهد. محاسبه این ضرایب را به کمک کمینه سازی رابطه زیر صورت دادیم:

$$f(W) = \| Y - XW \|_2^2 + \lambda \| W \|_1^1 \qquad (۱۲-۴)$$

جمله‌ی اول رابطه، مجموع مربعات خطا، و جمله‌ی دوم، رگولاریزاسیون نرم L_1 روی W هستند. λ پارامتر متعادل‌کننده این دو جمله است. هر مدل برآورد شده تنها با استفاده از حداقل

[۱] Regression coefficients

مربعات خطا، مقداری واریانس در خود دارد، یعنی این مدل به خوبی برای یک مجموعه داده متفاوت از داده‌های آموزشیِ خودش تعمیم داده نمی‌شود. رگولاریزاسیون، به طور قابل‌توجهی واریانس مدل را، بدون افزایش قابل‌توجه در بایاس آن، کاهش می‌دهد. بنابراین پارامتر تنظیم λ که در بالا توضیح داده شد استفاده می‌شود، تا تاثیر بر بایاس و واریانس را کنترل کند.

ابتدا پارامتر λ از طریق اعتبار سنجی متقابل ۱۰ لایه[۱] بهینه‌سازی می‌شود. به گونه‌ای که داده‌های آموزشی به ۱۰ قسمت مساوی تقسیم شده و هر بار ۹ قسمت جهت آموزش و ۱ قسمت باقی‌مانده جهت اعتبارسنجی مدل استفاده می‌شود. انتخاب پارامتر بهینه به ازای بیشینه کردن دقت اعتبار سنجی متقابل صورت می‌گیرد. در ادامه با استفاده از پارامتر کنترل‌کننده بهینه و کل داده‌های آموزشی، مجددا مدل را اصلاح کرده تا به تخمین نهایی ضرایب دست یابیم.

آموزش شبکه در ۳۵ دوره و در هر دوره به ازای نمونه‌های ۱۰۰تایی از تصاویر انجام می‌شود. به‌روزرسانی ضرایب نیز براساس روش گرادیان کاهشی[۲](GD) صورت می‌گیرد. به منظور جلوگیری از بیش‌برازش مدل رگرسیون از تکنیک حذف تصادفی[۳] با احتمال حذف ۰/۳ استفاده می‌شود. در واقع با ضرب معکوس ماتریس ضرایب مدل رگرسیونی در سیگنال‌های fMRI، بدون در نظرگرفتن مقدار خطا، کدهای نهان به صورت زیر به‌دست می‌آیند:

$$X = YW^T \qquad (۱۳-۴)$$

۴-۴-۲- مدل شبکه عصبی

دومین رویکرد پیشنهادی در جهت رسیدن به مدل نگاشت سیگنال‌های fMRI، استفاده از یک شبکه رمزگذار ساده است. پیشنهادی که به‌طور مشخص در این پژوهش مطرح شده است. اگر

[۱] Fold
[۲] Gradient Descent
[۳] Dropout

در نظر بگیریم ورودی این شبکه نمونه‌های سیگنال fMRI و خروجی شبکه بردارهای بیان‌گر کدهای نهان باشند، طبق رابطه زیر می‌توان سیگنال‌های fMRI را به کدهای نهان مربوط ساخت:

$$X = WY \qquad (۱۴-۴)$$

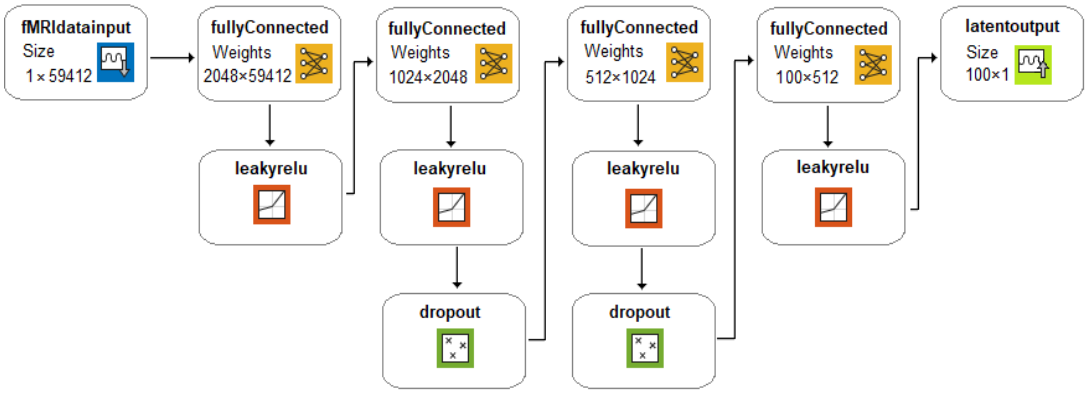
در رابطه ۴-۱۴، X بیان‌گر کدهای نهان مربوط به فریم‌های ویدئو تحریک است. ماتریسی $N \times l$ بُعدی، که در آن N تعداد حجم‌ها و l طول هر بردار نهان است. ماتریس Y، مربوط به سیگنال‌های fMRI است. این ماتریس ابعادی برابر $N \times v$ دارد که در آن v، بیان‌گر تعداد وکسل‌ها و N تعداد حجم‌ها است. W، ماتریس وزن شبکه است. این ماتریس دارای $l+1$ ستون و v سطر است.

اصلاح وزن‌های شبکه از طریق روش گرادیان کاهشی تصادفی و با کمینه سازی رابطه‌ی زیر انجام می‌شود:

$$f(W) = \| X - YW \|_2^2 \qquad (۱۵-۴)$$

رابطه ۴-۱۵ بیان‌گر آن است، که هر بار مربع اختلاف نمونه‌های کد نهان هدف، X، و نمونه‌های تولید شده توسط شبکه، YW، سنجیده می‌شود و براین اساس اصلاح وزن انجام می‌شود.

آموزش شبکه در ۳۵ دوره و در هر دوره به ازای نمونه‌های ۱۰۰تایی از ورودی انجام شد. به منظور جلوگیری از بیش‌برازش مدل شبکه عصبی نیز از تکنیک حذف تصادفی با احتمال حذف ۰/۲ استفاده می‌شود. ساختار شبکه به صورت نرونی و در ۴ لایه در نظر گرفته شده است. انتخاب چنین ساختاری به‌صورت تجربی حاصل شده است. جزییات ساختار در شکل ۴-۴ قابل مشاهده است.



شکل ۴-۴: معماری شبکه رمزگذار به‌کار گرفته شده جهت نگاشت سیگنال‌های fMRI به فضای نهان.

۴-۵- مدل رمزگشایی مغز

پس از برآورد نگاشتی مناسب که بتواند سیگنال‌های fMRI را به کدهای نهان مربوط سازد، با ترکیب این نگاشت و زیرشبکه رمزگشا آموزش دیده از قبل (بخش ۴-۳) به مدل رمزگشایی مدنظر دست می‌یابیم. و همانطور که در شکل ۴-۱ نشان داده شد، قادریم بازنمایی‌های قابل درکی از فریم‌های تحریک را، از روی داده‌های fMRI ارائه دهیم.

به این صورت که سیگنال‌های fMRI به عنوان ورودی به مدل نگاشت، داده می‌شوند، سپس در خروجی کدهای نهان مربوطه‌شان شکل می‌گیرد. با دادن این کدها به ورودی زیرشبکه رمزگشا، بازنمایی تصاویر حاصل می‌شود.

پس از دست‌یابی به مدل رمزگشاییِ مغزِ آموزش دیده، به منظور تست مدل، از داده‌های fMRI تست که از مشاهده فیلم ۵ بخشیِ تست، حاصل شده‌اند، استفاده می‌کنیم. جهت ارزیابی عملکرد مدل، شباهت بین بازنمایی‌های ایجاد شده توسط مدل و فریم‌های فیلم تست، براساس معیارهای معرفی شده در بخش ۴-۳ سنجیده می‌شود.

در اینجا نیز، به منظور ارزیابی مدل، نیاز داریم تا بخش‌های مختلف فیلم تست را فریم‌بندی کنیم. فریم‌بندی مشابه آنچه که در مراحل آموزش صورت گرفت، انجام می‌شود. به‌گونه‌ای که جهت استخراج فریم‌های فیلم ۴۰ دقیقه‌ای تست، ابعاد تصاویر، ۶۴ × ۶۴ پیکسل درنظر گرفته می‌شود. با

توجه به این‌که نرخ‌فریم برابر۳۰ فریم در ثانیه است، هر بخش ۸ دقیقه‌ای شامل ۱۴۴۰۱ تصویر می‌شود و به صورت کلی، به‌ازای همه ۵ بخش، نزدیک به ۷۲ هزار تصویر خواهیم داشت. به منظور مطابقت تعداد فریم‌ها با نرخ نمونه‌برداری fMRI (۲ هرتز)، فریم‌ها را با فرکانس ۳۰×۲ نمونه‌برداری کردیم. نهایتا به ازای ۴۰ دقیقه فیلم تست، ۱۲۰۰ فریم، تصویر طبیعی خواهیم داشت.

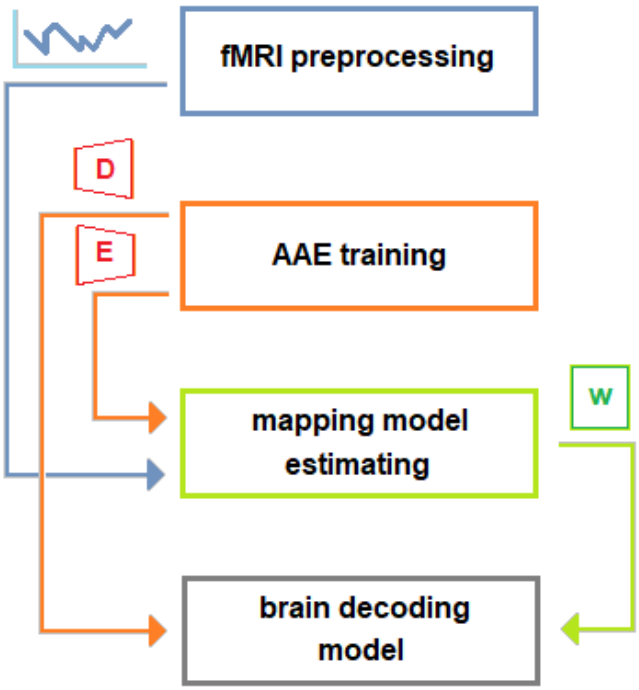
۴-۶- جمع‌بندی

با مطالعه این فصل از گزارشِ پژوهش انجام شده، گام‌های پیشنهادی جهت ارائه یک مدل رمزگشایی مغز، شناخته می‌شوند. مدلی که قادر به بازگردانی تصاویر قابل درکی از آن‌چه که انسان مشاهده کرده است، باشد.

گام‌های چهارگانه‌ای که به‌ترتیب عبارتند از:

۱. پیش‌پردازش داده‌های fMRI و استخراج سری‌های زمانی ناحیه بینایی، تحت عنوان سیگنال‌های fMRI.
۲. آموزش یک شبکه AAE با استفاده از تصاویر رویت نشده توسط سوژه‌ها، و بهره-گیری از زیرشبکه‌های رمزگذار و رمزگشای این ساختار، به‌ترتیب در گام‌های سوم و چهارم.
۳. آموزش نگاشتی جهت مرتبط کردن سیگنال‌های fMRI به کدهای نهان مربوط به فریم‌های تحریک، با به‌کارگیری دو رویکرد مبتنی‌بر مدل رگرسیون خطی چندمتغیره و مبتنی‌بر شبکه عصبی. جهت آموزش این نگاشت زیرشبکه رمزگذار آموزش دیده شده در گام دوم، به شکلی که در بخش ۴-۴ شرح داده شد، مورد استفاده قرار می‌گیرد.
۴. ایجاد مدل رمزگشایی مغز با ترکیب نگاشت آموزش دیده شده در گام سوم و زیرشبکه رمزگشا حاصل از آموزش مرحله دوم.

در شکل ۴-۵ روندنمای آن‌چه که تا به اینجا بیان شد، آمده است.



شکل ۴-۵: فلوچارت روند شکل‌گیری مدل رمزگشایی.

۵- فصل پنجم

ارائه یافته‌های پژوهش

۵-۱- مقدمه

در فصل قبل روش پیشنهادی این پژوهش، به منظور ارائه یک مدل رمزگشایی مغز، جهت بازسازی تصاویر درک شده توسط کاربر از روی داده‌های fMRI شرح داده شد. در فصل پیشِ‌رو، یافته‌های هر مرحله از روش پیشنهاد شده، در بخش‌های مجزا ارائه می‌شوند و به تحلیل نتایج به‌دست آمده می‌پردازیم.

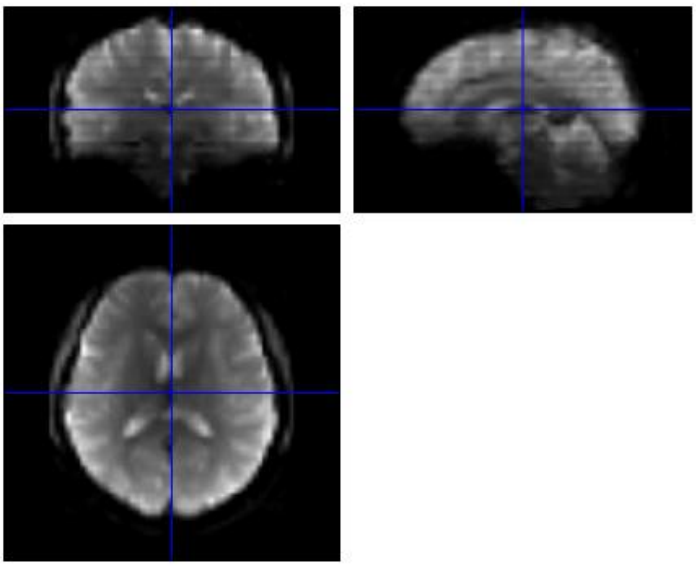
۵-۲- پیش‌پردازش داده‌ها

پیش‌پردازش داده‌ها در نرم‌افزار SPM12 انجام شد. مراحل انجام شده در محیط نرم‌افزار به شرح زیر است:

۱. "Slice timing": با انجام این مرحله اختلاف زمان‌های بُرش‌های مغز اصلاح می‌شود. در تمام داده‌ها، تعداد بُرش‌ها برابر ۳۸ بُرش، زمان تکرار، TR، برابر ۲ ثانیه و زمان بین دسترسی اولین بُرش و آخرین بُرش، TA، نیز براساس رابطه ۵-۱ قابل محاسبه است.

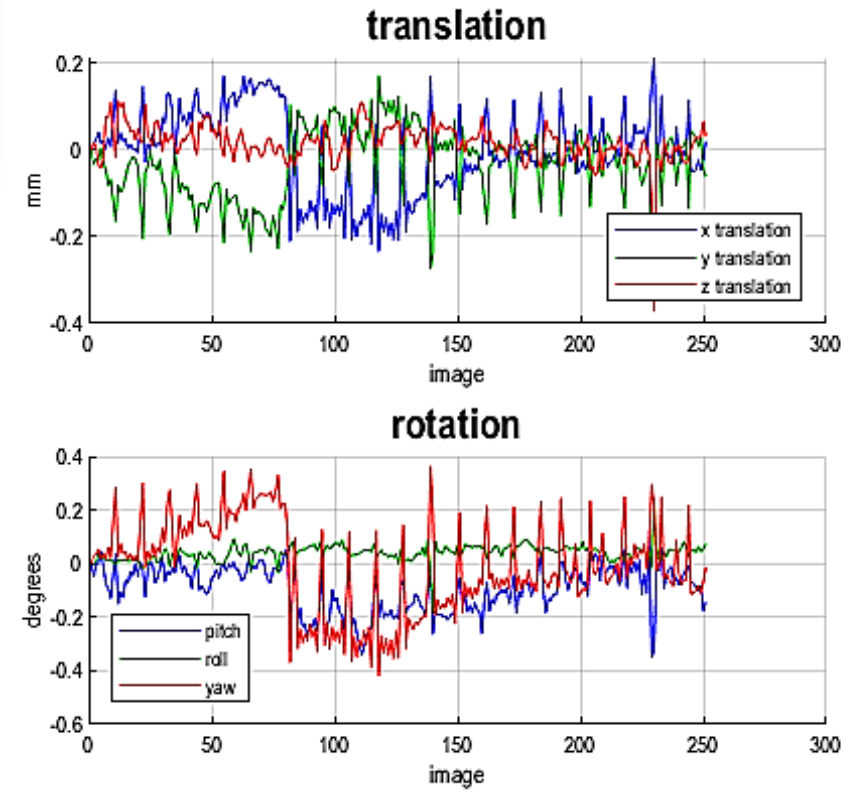
$$TA = TR - \left(TR / Number\ of\ Slices\right) \qquad (۱-۵)$$

مدل دست‌یابی به بُرش‌ها نیز به عنوان ورودی این بخش مورد نیاز است، که برای تمام داده‌ها از روش Interleaved جهت اخذ داده‌ها استفاده شده است. خروجی این مرحله به ازای یک نمونه داده در شکل ۵-۱ نمایش داده شده است.



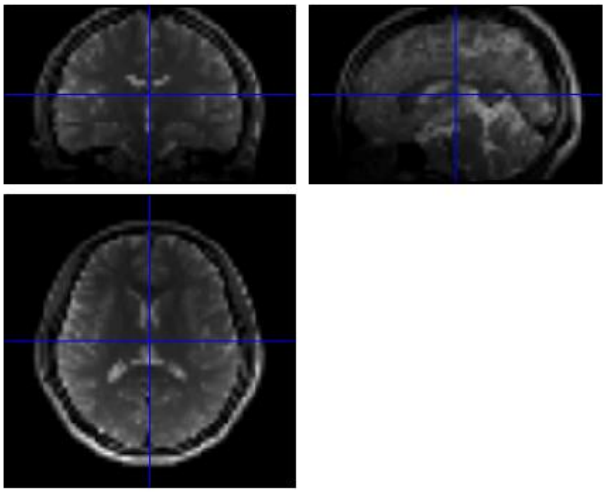
شکل ۵-۱: خروجی مرحله Slice timing بر روی داده مربوط به مشاهده اولین بخش ویدئو آموزشی توسط سوژه اول.

۲. "Realign": این تابع ساده‌ترین تابع برای تطبیق ثبت‌های مختلف است. در واقع این تابع فقط حرکت تصویر در راستای محورهای X، Y و Z و دوران حول این محورها را انجام می‌دهد. هدف این تابع کمینه کردن اختلاف دو تصویر با روش آزمون و خطاست. تابع هزینه‌ای که کمینه می‌شود، مربعات اختلاف دو تصویر است. به طور کلی، این تابع به‌منظور اصلاح حرکت سر کاربرد دارد. خروجی این مرحله به ازای یک نمونه داده در شکل ۵-۲ نمایش داده شده است.



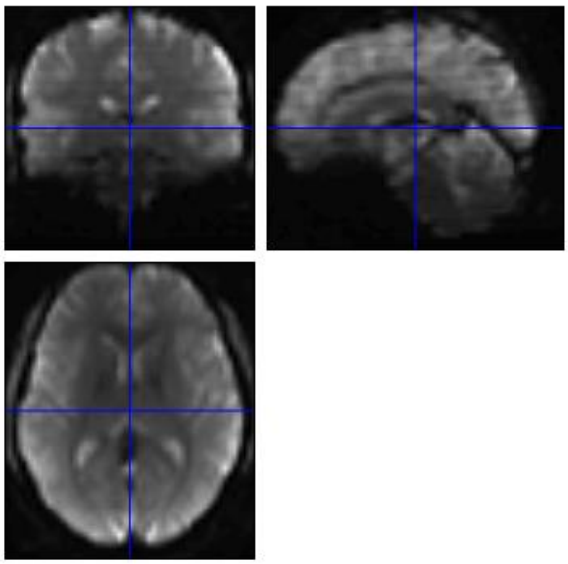
شکل ۵-۲: خروجی مرحله Realign بر روی داده مربوط به مشاهده اولین بخش ویدئو آموزشی توسط سوژه اول. این داده شامل ۲۵۱ حجم متوالی بوده و محدوده تغییرات در انتقال و چرخش در تصویر قابل مشاهده است.

۳. "Coregister": این مرحله جهت تطبیق تصاویری که مُدالیته و ساختارهای متفاوتی دارند انجام می‌شود. در اینجا برای همسان کردن دو تصویر با وزن‌دهی‌های متفاوت T_1 و T_2، می‌توان از این گزینه استفاده کرد. به عبارتی به کمک این تابع تصویر T_1 از فضای ساختاری به فضای کارکردی منتقل می‌شود. خروجی این مرحله به ازای یک نمونه داده در شکل ۵-۳ نمایش داده شده است.



شکل ۵-۳: خروجی مرحله Coregister بر روی داده مربوط به مشاهده اولین بخش ویدئو آموزشی توسط سوژه اول.

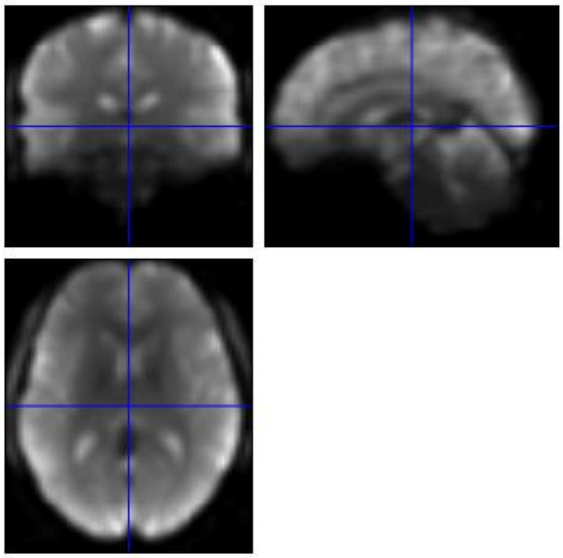
۴. "Normalise": در این مرحله تمام تصاویر کارکردی به فضای استاندارد منتقل می‌شوند. خروجی این مرحله به ازای یک نمونه داده در شکل ۵-۴ نمایش داده شده است.



شکل ۵-۴: خروجی مرحله Normalise بر روی داده مربوط به مشاهده اولین بخش ویدئو آموزشی توسط سوژه اول.

۱. "Smoothing": این تابع به عنوان آخرین مرحله از پیش‌پردازش برای هموار کردن تصاویر استفاده می‌شود. البته با اعمال این تابع وضوح تصویر کاهش می‌یابد. لذا

میزان هموار کردن بسته به هدف پردازش متغیر است. خروجی این مرحله به ازای یک نمونه داده در شکل ۵-۵ نمایش داده شده است.



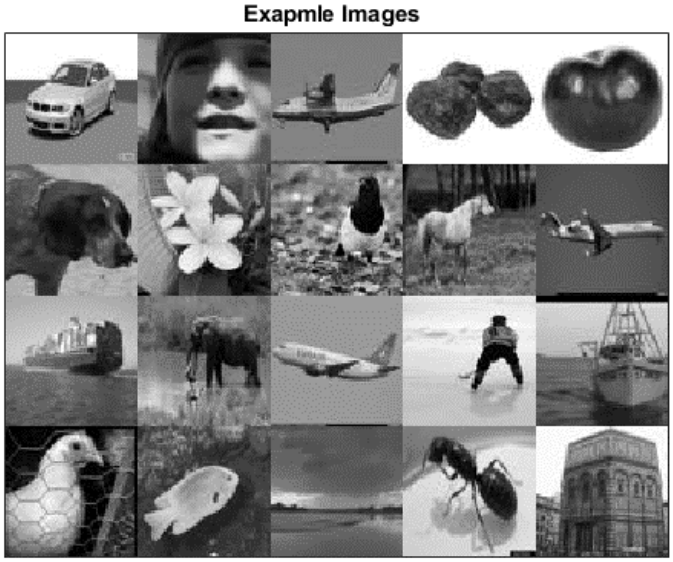
شکل ۵-۵: خروجی مرحله Smoothing بر روی داده مربوط به مشاهده اولین بخش ویدئو آموزشی توسط سوژه اول.

۵-۳- آموزش و تست شبکه خودرمزگذار رقابتی

در بخش ۴-۳، به صورت مختصر ساختار شبکه پیشنهادی و نحوه به‌کارگیری آن شرح داده شد. در اینجا هدف آن است که، روند آموزش و تست شبکه به تفصیل بیان شده، و نتایج به‌دست آمده ارائه شوند. بر این اساس قادر خواهیم بود تحلیل مناسبی از نتایج حاصل شده، ارائه دهیم.

در ابتدا جهت آموزش شبکه خودرمزگذار رقابتی، یک مجموعه ۱۱۲۵۰تایی از تصاویر طبیعی از مجموعه تصاویر ImageNet و COCO نمونه‌برداری شدند. این مجموعه به ۱۵ گروه کلیِ موجود در فیلم‌های آموزش و تست، تحت عناوین؛ محیط سرپوشیده، محیط سرباز، انسان، چهره، پرنده، حشره، حیوانات آب‌زی، حیوانات خشکی‌زی، گل، میوه، چشم‌اندازهای طبیعی، ماشین، هواپیما، کشتی و ورزش کردن، تقسیم‌بندی شده، و به‌صورت دستی و توسط انسان برچسب گذاری شدند. تمام گروه‌ها تعداد تصاویر برابری را شامل می‌شوند، بنابراین در هر گروه ۷۵۰ تصویر طبیعی خواهیم داشت. این گروه‌بندی‌ها محتوای مشترکی را در هر دو فیلم‌های آموزش و تست، پوشش

می‌دهند. پس از آن نیز بررسی بصری مجددی به‌منظور جایگزینی تصاویر دارای برچسب اشتباه، انجام شد. سپس تمامی تصاویر این مجموعه در سطح خاکستری[۱] و با ابعاد ۶۴ × ۶۴ پیکسل ذخیره شدند تا به عنوان ورودیِ شبکه مورد استفاده قرار گیرند. نمونه‌هایی از مجموعه تصاویر آموزشی در شکل ۵-۶ نمایش داده شده است.



شکل ۵-۶: نمونه‌هایی از مجموعه تصاویر آموزشی جهت آموزش شبکه خودرمزگذار رقابتی.

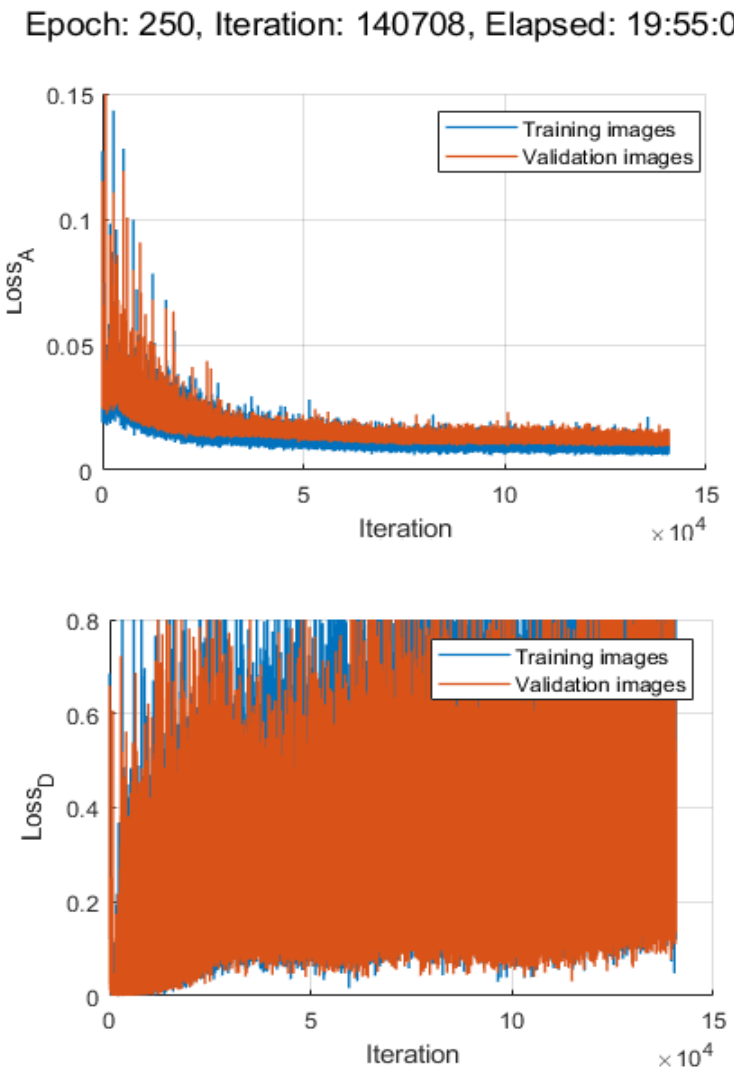
در ابتدا ضرایب وزنی هر زیرشبکه از این ساختار با یک توزیع گاوسی تصادفی، با میانگین صفر و واریانس مشخص مقداردهی اولیه می‌شوند: $W \sim \mathcal{N}(\circ,\sigma)$. سپس مجموعه تصاویر آماده شده را به صورت تصادفی بُر زده و یک زیرمجموعه ۲۰تایی از آن را به عنوان داده‌های ارزیابی جدا نمودیم. در ادامه آموزش شبکه در ۲۵۰ دوره و در هر دوره به ازای نمونه‌های ۲۰تایی از ۱۱۲۳۰ تصویر باقی‌مانده، انجام می‌شود. بدین صورت که، در هر دوره، مجموعه آموزشی به‌صورت تصادفی به ۵۶۱ بسته ۲۰تایی از تصاویر تقسیم می‌شود و اصلاح ضرایب وزنی شبکه به‌ازای هر دسته و براساس روابط ۴-۱ و ۴-۲، صورت می‌گیرد. آمورش شبکه برپایه واحد پردازش مرکزی[۲] (CPU)، با

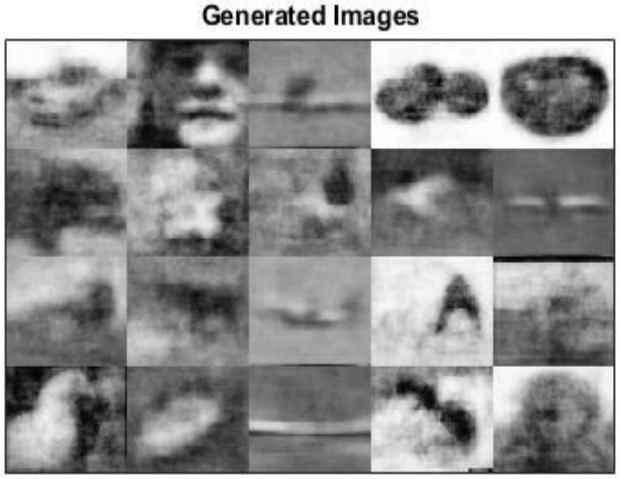
[۱] Gray level

[۲] Central Processing Unit

استفاده از یک رایانه دارای واحد پردازش مرکزی Core i3 مدل ۷۱۰۰ و فرکانس پردازشی ۳/۹۰ گیگاهرتز، حدود ۲۰ ساعت به‌طول انجامید.

در حین آموزش شبکه، به ازای هر تکرار، مقادیر خطا زیرشبکه‌های تفکیک‌کننده و رمزگذار گزارش شده است. به ازای هر ۲۰ تکرار نیز، ۲۰ تصویر ارزیابی به‌عنوان ورودی به شبکه اعمال می‌شوند و تصاویر تولید شده توسط شبکه در خروجی نمایش داده می‌شوند. براساس این تصاویر نمایش داده شده، شبکه در دوره ۲۵۰ام تصاویر نسبتا قابل درکی را - براساس درک بینایی انسان - تولید نمود، بنابراین فرآیند آموزش را در این دوره متوقف نمودیم (شکل ۵-۷).

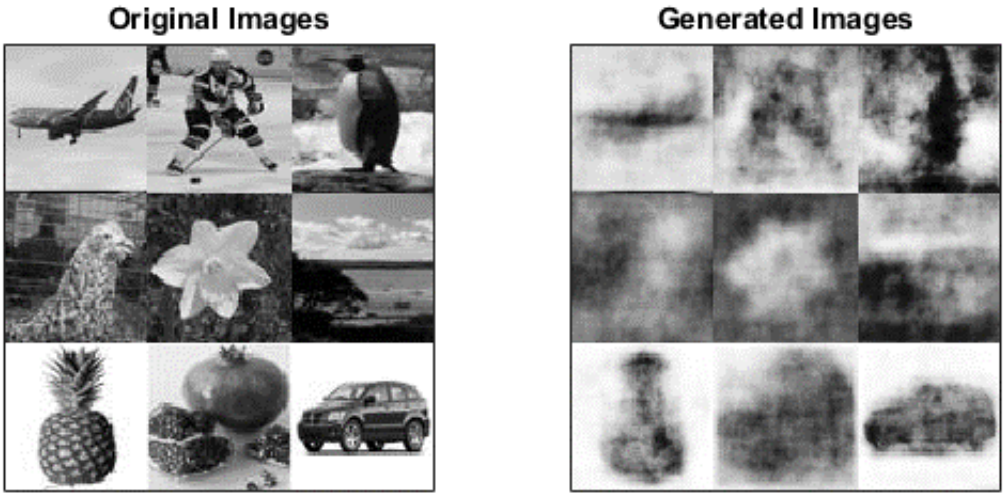




شکل ۵-۷: گزارش روند آموزش شبکه خودرمزگذار رقابتی. (۱) تغییرات خطا به ازای زیرشبکه‌های رمزگذار و رمزگشا. (۲) تغییرات خطا به ازای زیرشبکه تفکیک‌کننده. (۳) تصاویر تولید شده در خروجی شبکه خودرمزگذار رقابتی به‌ازای تصاویر ارزیابی.

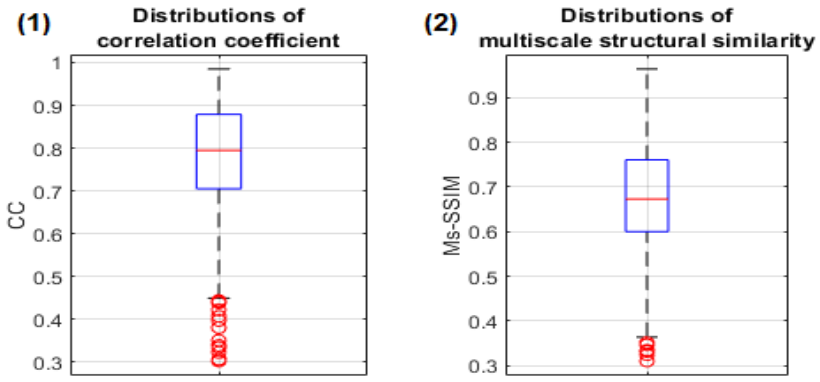
در بخش اول از شکل ۵-۷ نمودار Loss_A یا همان مقدار خطا در زیرشبکه‌های رمزگذار و رمزگشا، به ازای هر تکرار به تصویر کشیده شده است. بخش دوم نیز نمایش دهنده Loss_D، مقدار خطا در زیرشبکه تفکیک‌کننده، به ازای هر تکرار است. این دو مقدار به‌ازای داده‌های آموزشی (آبی) و ارزیابی (نارنجی) به صورت جداگانه ثبت شده‌اند. در نهایت در بخش چهارم از این شکل، تصاویر تولید شده توسط شبکه به‌ازای تصاویر ارزیابی را می‌توان مشاهده نمود. توقف آموزش شبکه براساس بررسی کیفی تصاویر تولید شده از روی تصاویر ارزیابی، انجام گرفت.

به‌منظور ارزیابی عملکرد شبکه، از مجموعه تصاویری شامل ۱۵۰۰ تصویر، استفاده کردیم. در این مجموعه از هر گروه ۱۰۰ تصویر، مشابه آن‌چه که برای تصاویر آموزشی بیان شد، جمع‌آوری شده است. شایان ذکر است، مجموعه تصاویر آموزشی و تست، متفاوتند و در مرحله تست، زیرشبکه تفکیک‌کننده کنار گذاشته شده و تصاویر تنها از دو زیرشبکه رمزگذار و رمزگشا عبور داده شدند. این تصاویر به‌صورت کاملا تصادفی به‌عنوان ورودی به شبکه اعمال شدند. نمونه‌هایی از تصاویر تولید شده در خروجیِ این شبکه، در شکل ۵-۸ نمایش داده شده است. به‌طوری‌که در سمت چپ، تصاویر اصلی و سمت راست، تصاویر تولید شده توسط شبکه قابل مشاهده هستند. همانطور که دیده می‌شود، بازنمایی‌ها تا حد زیادی قابل درک توسط انسان هستند.



شکل ۵-۸: تصاویر تولید شده توسط شبکه خودرمزگذار رقابتی به ازای تعدادی از تصاویر مجموعه تست.

جهت ارائه معیاری کمی در زمینه ارزیابی عملکرد شبکه، دو کمیت شباهت ساختاری و ضریب همبستگی میان تصاویر تستِ ورودی و تصاویر بازسازی شدهِ خروجی، سنجیده شده است. به‌منظور ارائه نمایشی گویا از مقادیر بدست آمده، توزیع هر کدام از معیارها به تفصیل آماری در شکل ۵-۹ به‌صورت نمایش جعبه‌ای بررسی شده است.



شکل ۵-۹: نمایش جعبه‌ای توزیع معیارهای شباهت. (۱) نمایش جعبه‌ای توزیع CC محاسبه شده. (۲) نمایش جعبه‌ای توزیع Ms-SSIM محاسبه شده.

براساس شکل ۵-۹ قسمت اول، می‌توان گفت، توزیع ضرایب به‌دست آمده میانه‌ای برابر با ۰/۷۹ داشته، به این معنی که، نیمی از تصاویر دارای ضرایب کوچک‌تر و نیمی دیگر دارای ضرایب بزرگ‌تر از این مقدار هستند. همچنین ۲۵ درصد از تصاویر ضرایبی کمتر از ۰/۷۱، و بیشتر از ۰/۸۸ داشته‌اند. بیشترین و کمترین مقدار حاصل شده به ترتیب عبارت‌اند از؛ ۰/۹۸ و ۰/۴۴. از میان ۱۵۰۰

تصویر در نظر گرفته شده، ۱۳ تصویر مقدار ضریب همبستگی پَرت، نزدیک به مقدار حداقل داشته‌اند. تصاویری که شاید بتوان گفت شبکه قادر به بازسازی آن‌ها نبوده است. همچنین درخصوص شباهت ساختاری براساس شکل ۵-۱۰ قسمت دوم، می‌توان گفت، توزیع مقادیر به‌دست آمده میانه‌ای برابر با ۰/۶۷ داشته است. بیشترین و کمترین مقدار حاصل شده به ترتیب عبارت‌اند از؛ ۰/۹۶ و ۰/۳۱. از میان ۱۵۰۰ تصویر در نظر گرفته شده، ۷ تصویر مقدار ضریب همبستگی پَرت، نزدیک به مقدار حداقل داشته‌اند.

بنابر نتایج به‌دست آمده می‌توان گفت، ساختار و روند آموزشی درنظر گرفته شده برای شبکه خودرمزگذار رقابتی، مطلوب بوده است و این شبکه‌ی آموزش دیده، قابلیت استفاده در سایر مراحل پژوهش را داراست.

۵-۴- آموزش و تست مدل نگاشت fMRI

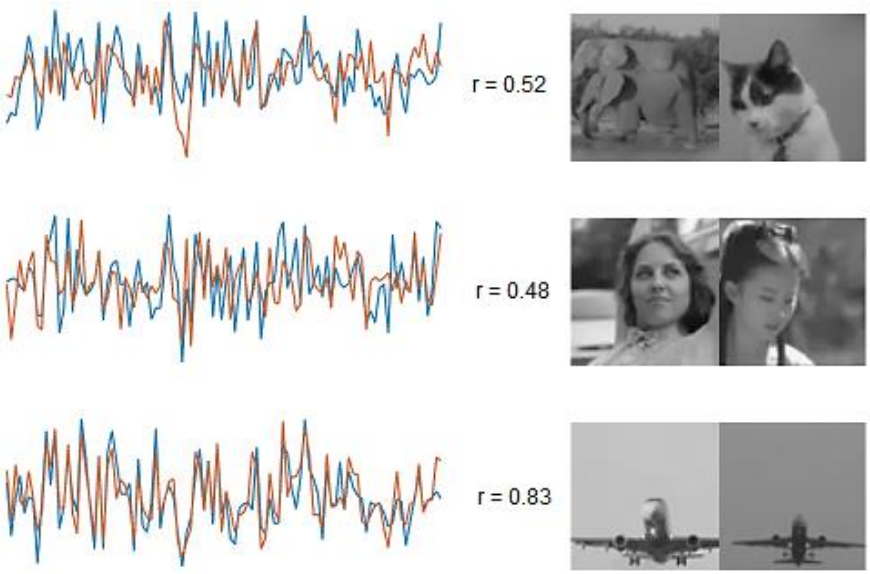
همان‌طور که در بخش ۴-۴ نیز بیان شد، این نگاشت با استفاده از دو رویکرد رگرسیون خطی چندمتغیره و شبکه عصبی برآورد شده است. در این بخش برآنیم که، روند آموزش و تست هر یک از رویکردهای پیشنهادی را به تفصیل بیان کنیم. در ادامه نتایج هر دو رویکرد به‌صورت جداگانه ارائه می‌شوند و آن‌ها را مورد بررسی قرار می‌دهیم.

۵-۴-۱- آموزش و تست مدل رگرسیون خطی چندمتغیره

در بخش ۴-۴-۱ رابطه حاکم بر مدل رگرسیون خطی چندمتغیره در قالب رابطه ۴-۱۱ معرفی شد. دانستیم که در مرحله آموزشِ مدل، ماتریس طراحی، x ، همان کدهای نهان مربوط به فریم‌های فیلم آموزشی است. به منظور ایجاد این ماتریس ابتدا لازم است فریم‌های فیلم آموزشی استخراج شوند. فریم‌بندی فیلم آموزشی ۲/۴ ساعته، به منظور مطابقت تعداد فریم‌ها با نرخ نمونه-برداری fMRI یا به‌عبارتی تعداد حجم‌های ثبت شده، با فرکانسِ نمونه‌برداری ۶۰ انجام شد. ابعاد تصاویر برابر با ابعاد تصاویر ورودی در شبکه خودرمزگذار رقابتی (۶۴ × ۶۴ پیکسل) درنظر گرفته شد و تصاویر به صورت خاکستری ذخیره شدند. نهایتا به ازای ۲/۴ ساعت فیلم آموزشی، ۴۳۲۱

فریم، تصویر طبیعی خواهیم داشت. سپس با عبور هر ۴۳۲۱ فریم، از زیرشبکه رمزگذار، آموزش دیده در قبل، کدهای نهان مربوط به این تصاویر ایجاد شدند. بنابراین پس از این فرآیند، به ماتریس طراحی ۱۰۱×۴۳۲۱ بُعدیِ مدنظر دست می‌یابیم. ۱۰۰ ستون اولیه از این ماتریس اشاره به کدهای نهان محاسبه شده دارند و آخرین ستون از این ماتریس، برداری است که مقادیر تمام عناصر آن برابر یک و به منظور اعمال مقدار ثابت در مدل رگرسیون، در نظر گرفته شده است.

قبل از آموزش نگاشت با استفاده از ماتریس طراحی به‌دست آمده، نشان می‌دهیم که زیرشبکه رمزگذار به ازای تصاویر مختلف از گروه‌های یکسان، کدهای نهان نسبتا همبسته‌ای تولید می‌کند. شکل ۵-۱۰ نیز گویای این مسئله است، به‌طوری که در آن کدهای نهان ایجاد شده را به ازای چند نمونه تصویر از چند گروه کلاسیِ درنظر گرفته شده، به تصویر کشیده است.

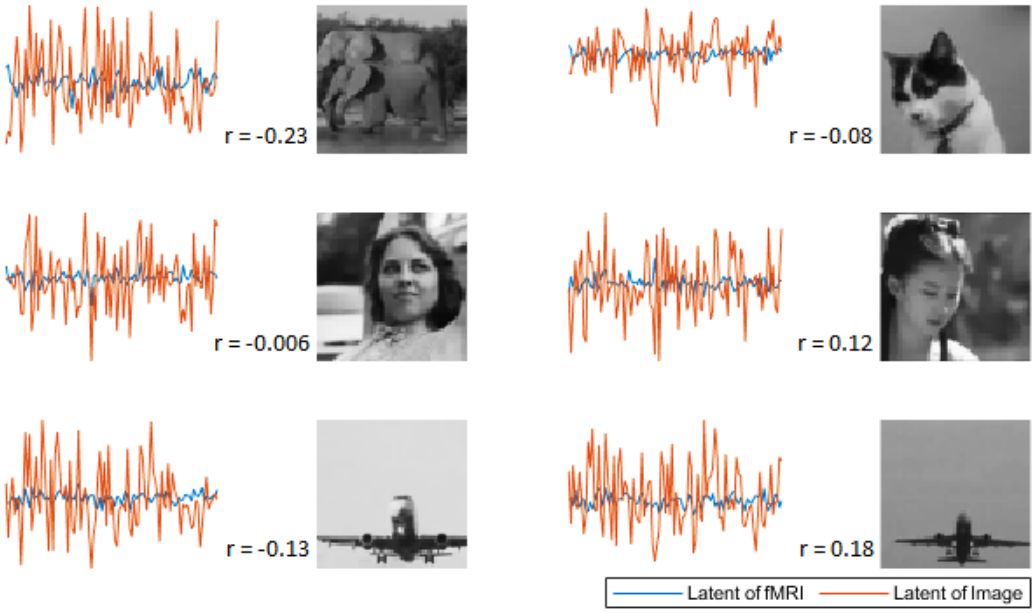


شکل ۵-۱۰: کدهای نهان ایجاد شده به ازای دو تصویر از سه کلاس متفاوت.

ماتریس Y، متغیر پاسخ نیز، مربوط به سیگنال‌های fMRI است. بنابراین ابعادی برابر ۴۳۲۱×۵۹۴۱۲ دارد. در اینجا تنها از سیگنال‌های fMRI یکی از سه سوژه انسانی به‌منظور آموزش مدل بهره می‌گیریم. هدف محاسبه ماتریس ضرایب، W، است و محاسبه این ضرایب به کمک کمینه-سازی رابطه ۴-۱۲، بیان شده در بخش ۴-۴-۱ صورت می‌گیرد.

ابتدا پارامتر λ از طریق اعتبار سنجی متقابل ۱۰ لایه بهینه‌سازی شد. به گونه‌ای که داده‌های آموزشی به ۱۰ قسمت مساوی تقسیم شده و هر بار ۹ قسمت جهت آموزش و ۱ قسمت باقی‌مانده جهت اعتبارسنجی مدل استفاده می‌شد. انتخاب پارامتر بهینه به ازای بیشینه کردن دقت اعتبار سنجی متقابل - همبستگی میان کدهای نهان حاصل از مدل با کدهای نهان ایجاد شده - صورت گرفت. در ادامه با استفاده از پارامتر کنترل‌کننده بهینه و کل داده‌های آموزشی، مجددا مدل را اصلاح کرده تا به تخمین نهایی ضرایب دست یابیم. به‌روزرسانی ضرایب در این روند آموزشی، براساس رویکرد گرادیان کاهشی انجام می‌شود.

به‌منظور تست نگاشت، از مجموعه سیگنال‌های fMRI یکی از دو سوژه دیگر استفاده نمودیم. بررسی عملکرد مدل، براساس محاسبه همبستگی میان کدهای نهان حاصل از مدل با کدهای نهان ایجاد شده از روی تصاویر ویدئویی، انجام شد. این بررسی در قالب شکل ۵-۱۱ به تصویر کشیده شده است.



شکل ۵-۱۱: بررسی عملکرد مدل نگاشت fMRI، مبتنی‌بر رگرسیون خطی چندمتغیره به ازای داده‌های آموزشی سوژه دوم.

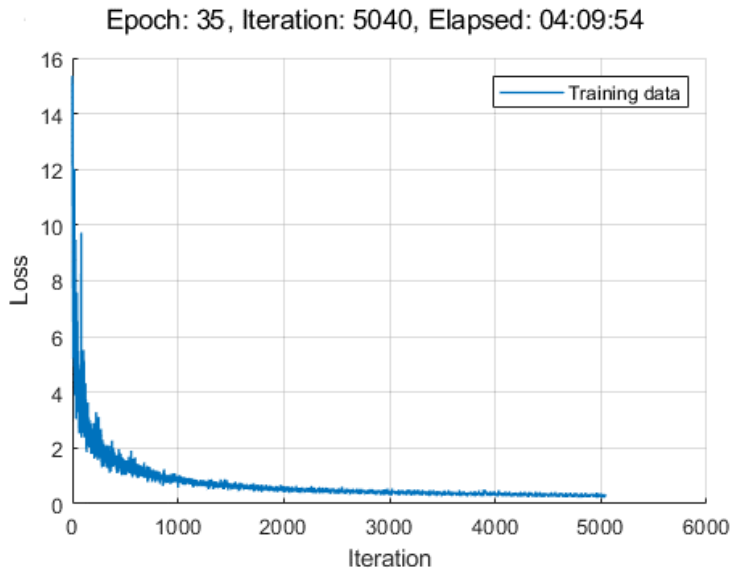
آن‌چه که از تصویر ۵-۱۱ قابل درک است، این است که، مدل نگاشت برپایه رگرسیون خطی نتوانسته است به‌درستی داده‌های fMRI را به کدهای نهان تصاویر تبدیل کند. به عبارتی هدف اصلی از این نگاشت، یعنی انتخاب وکسل‌های برانگیخته شده به ازای مشاهده تصاویر متفاوت از میان ۵۹۴۱۲ وکسل موجود، محقق نشده است. البته قابل ذکر است که سنجش اصلی این مدل با بررسی توانمندی آن در بازیابی تصاویر تحریک، در ادامه محقق می‌شود.

۵-۴-۲- آموزش و تست مدل شبکه عصبی

براساس رابطه ۴-۱۴، بیان شده در بخش ۴-۴-۲، سیگنال‌های fMRI، Y، به کدهای نهان مربوط به فریم‌های ویدئو تحریک، X، مرتبط می‌شوند. سیگنال‌های fMRI در طی آموزش، ابعادی برابر ۵۹۴۱۲×۴۳۲۱ داشت. کدهای نهان مربوط به فریم‌های فیلم آموزشی نیز مشابه آن‌چه در بخش ۴-۴-۱ بیان شد، مورد استفاده قرار گرفت با این تفاوت که، در اینجا ستون مربوط به جمله‌ی ثابت درنظر گرفته نمی‌شود. بنابراین این ماتریس نیز ابعادی برابر ۱۰۰×۴۳۲۱ دارد. در آموزش شبکه، ماتریس سیگنال‌های fMRI به عنوان ورودی به شبکه اعمال می‌شود یعنی هر حجم از ثبت fMRI، در ورودی به‌عنوان یک نمونه درنظر گرفته می‌شود. همچنین هر نمونه - هر فریم ویدئویی - از ماتریس کدهای نهان به عنوان برچسبی برای هر نمونه از ورودی، مورد استفاده قرار می‌گیرد. هدف محاسبه ماتریس ضرایب وزنی، W، است. ماتریسی که نگاشت رابطه ۴-۱۴، بیان شده در بخش ۴-۴-۲، را محقق می‌سازد.

در ابتدای آموزش، مقادیر ضرایب به صورت تصادفی و براساس یک توزیع گاوسی، با میانگین صفر و واریانس مشخص درنظر گرفته می‌شوند: $W \sim \mathbb{N}(\circ,\sigma)$. در هر دوره از ۳۵ دورهِ آموزشی، نمونه‌ها به صورت تصادفی به بسته‌های ۱۰۰تایی تقسیم‌بندی می‌شوند. و هر بار به روزرسانی ضرایب ماتریس وزنی براساس تکنیک گرادیان کاهشی تصادفی، روی یک بسته انجام می‌شود. به منظور جلوگیری از بیش‌برازش مدلِ شبکه عصبی نیز، از تکنیک حذف تصادفی با احتمال حذف ۰/۳

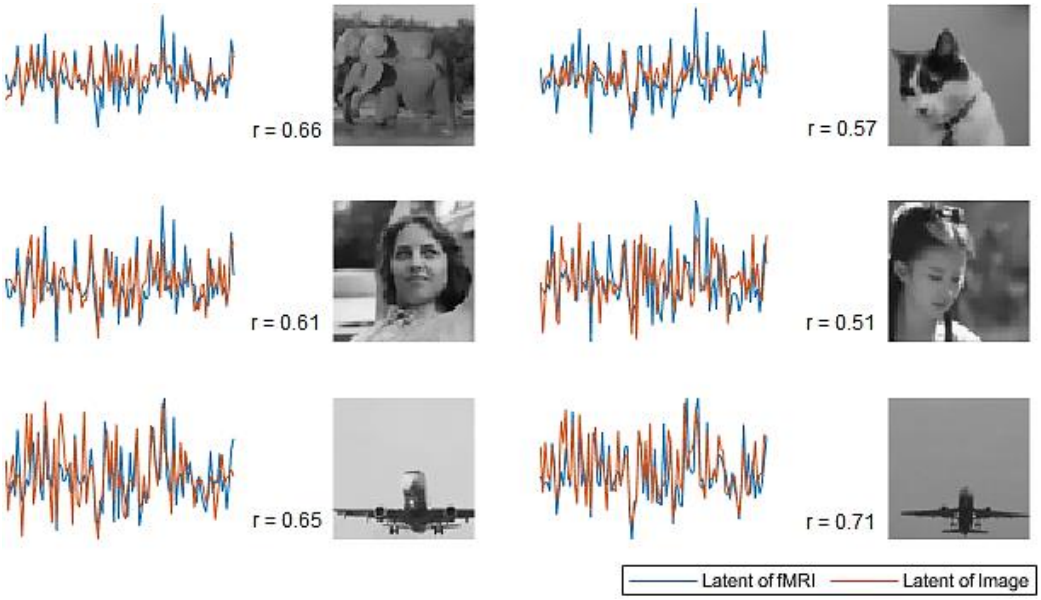
استفاده کردیم. مقدار خطا در حین آموزش در شکل ۵-۱۲ به ازای داده‌های آموزشی نشان داده شده است.



شکل ۵-۱۲: نمایش مقدار خطا در حین آموزش شبکه رمزگذار، به ازای داده‌های آموزشی.

بررسی عملکرد مدل آموزش دیده نیز مشابه آن‌چه که برای مدل رگرسیون خطی گفته شد، صورت گرفت. در اینجا نیز باید گفت عملکرد مدل، زمانی به درستی سنجیده می‌شود که در ترکیب رمزگشای مغز قرار گیرد.

در شکل ۵-۱۳ نتایج بررسی اولیه از عملکرد مدل قابل مشاهده است. به نحوی که در آن به ازای چند حجم متفاوت، داده‌های fMRI آموزشی از سوژه دوم، با استفاده از مدل رمزگذار آموزش داده در اینجا، به کدهای نهان تبدیل شده‌اند. در ادامه همبستگی میان آن‌ها و کدهای نهان مربوط به تصاویر متناظر با این حجم‌ها سنجیده شده است.



شکل ۵-۱۳: بررسی عملکرد مدل نگاشت fMRI، مبتنی‌بر شبکه عصبی به ازای داده‌های آموزشی سوژه دوم.

براساس نتایج حاصل شده، می‌توان گفت، مدل نگاشت مبتنی‌بر شبکه عصبی به شکل بهتری نسبت به مدل مبتنی‌بر رگرسیون خطی چندمتغیره، توانسته از میان تمام وکسل‌ها، وکسل-های برانگیخته شده متناظر با هر تصویر تحریک را انتخاب کند. از این‌رو انتظار بر آن است که، به‌کارگیری این مدل در ادامه روند پژوهش نیز نتایج مطلوب‌تری داشته باشد.

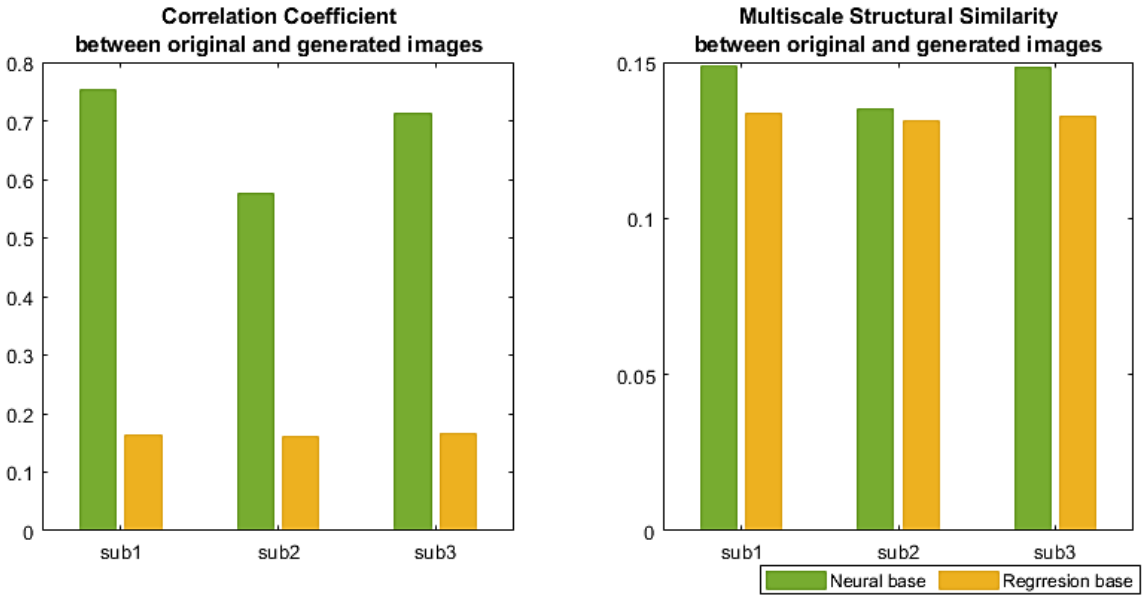
۵-۵- تست مدل رمزگشایی مغز

پس از برآورد مدل نگاشت و آموزش زیرشبکه رمزگشا با ترکیب این دو ساختار به مدل نهایی رمزگشایی مغز دست می‌یابیم. هدف آن است که بتوانیم با استفاده از مدل آموزش دیده، داده‌های fMRI به ازای سوژه‌های مختلف را رمزگشایی نماییم. براساس آن‌چه که در قبل نیز بیان شد، آموزش مدل رمزگشایی تنها با به‌کارگیری داده‌های fMRI آموزشی یکی از سه سوژه انجام گرفت.

در ادامه، جهت ارزیابی، داده‌های fMRI مربوط به مشاهده ویدئوی تست، مورد استفاده قرارمی‌گیرند. به‌گونه‌ای که سیگنال‌های fMRI از هر یک از سوژه‌ها به ورودی مدل داده می‌شود و

در خروجی فریم‌های بازسازی شده‌ی فیلم تست را خواهیم داشت. فریم‌بندی فیلم تست نیز مشابه آن‌چه که برای فیلم آموزشی بیان شد، صورت گرفت.

به منظور سنجش عملکرد، دو معیار ضریب همبستگی و شباهت ساختاری چندمقیاسه بین تصاویر بازسازی شده به ازای داده‌های تستِ هر یک از سه سوژه، و فریم‌های ویدئوی تست، در هر دو مدلِ درنظر گرفته شده، محاسبه شده است. مقادیر به‌دست آمده به ازای تمام فریم‌ها میانگین-گیری شده و در شکل ۵-۱۴ قابل مشاهده هستند.



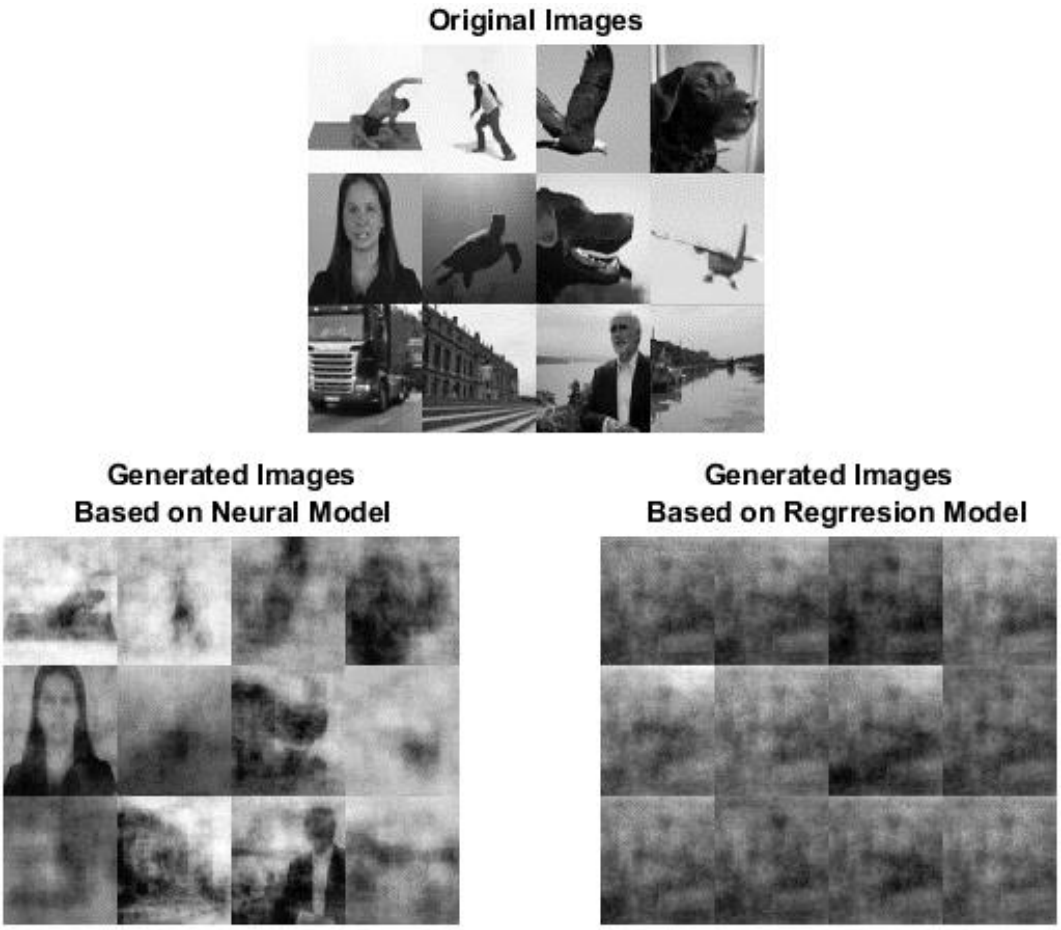
شکل ۵-۱۴: نتایج ارزیابی کمی مدل رمزگشایی مغز. در سمت راست، میانگین معیار شباهت ساختاری چندمقیاسه به ازای هر سوژه و دو مدل درنظر گرفته شده، نمایش داده شده است. در سمت چپ، میانگین معیار ضریب همبستگی به ازای هر سوژه و دو مدل درنظر گرفته شده، نمایش داده شده است.

همانطور که انتظار می‌رفت، براساس نتایج قبلیِ به‌دست آمده برای مدل مبتنی‌بر رگرسیون خطی چندمتغیره، در اینجا نیز نتایج قابل قبولی از این مدل نداریم. اما مدل مبتنی‌بر شبکه عصبی عملکرد چشم‌گیری دارد.

با مقایسه معیارهای کمیِ محاسبه شده، می‌توان گفت معیار شباهت ساختاری چندمقیاسه به علت حساسیت بالا نسبت به نویز در تصاویر، معیار چندان مناسبی به منظور ارزیابی کمی چنین

مدل‌هایی محسوب نمی‌شود. چرا که براساس نتایج این معیار، تفاوت معناداری میان عملکرد دو مدل پیاده‌سازی شده، وجود ندارد.

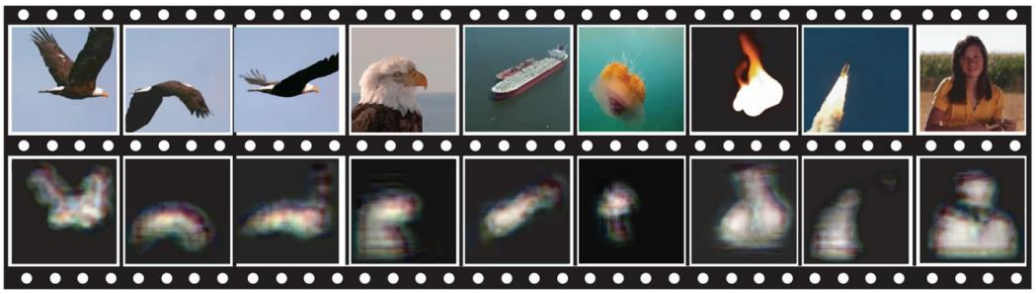
در شکل ۵-۱۵ نمونه تصاویر بازسازی شده به ازای هر یک از مدل‌ها در مقایسه با تصاویر اصلی قابل مشاهده است.



شکل ۵-۱۵: نتایج کیفی مدل رمزگشایی مغز. در قسمت بالایی چند نمونه از تصاویر اصلی فریم‌های ویدئو تست نمایش داده شده است. در قسمت پایین سمت راست، تصاویر بازسازی شده به ازای داده‌های تست سوژه اول، با استتفاده از مدل مبتنی‌بر رگرسیون خطی چندمتغیره، نمایش داده شده است. در قسمت پایین سمت چپ، تصاویر بازسازی شده به ازای داده‌های تست سوژه اول، با استتفاده از مدل مبتنی‌بر شبکه عصبی، نمایش داده شده است.

در نهایت براساس مشاهده نتایج کمی و کیفی به‌دست‌آمده، می‌توان گفت مدل رمزگشایی مغز پیشنهاد شده در این پژوهش، عملکرد چشم‌گیری در زمینه رمزگشایی داده‌های fMRI به ازای مشاهده تصاویر طبیعی، داشته است.

و اما در رابطه با دو مطالعه‌ی اصلی [۹۲, ۹۳] که هر کدام به‌عنوانی الهام‌بخش شکل‌گیری این پژوهش بوده‌اند باید گفت؛ بهره‌گیری از اولین مطالعه [۹۲]، به‌منظور استفاده از مجموعه داده-های fMRI معرفی شده در آن بوده است. در این مطالعه روش پیشنهادی برپایه شبکه‌های همگستی عمیق و رگرسیون خطی، به شرح زیر بوده است. ابتدا ضرایب وزنی یک شبکه همگشتی عمیق از پیش آموزش دیده شده، AlexNet، با استفاده از مجموعه تصاویر رویت نشده توسط سوژه‌های انسانی و شامل ۱۵ کلاس طبیعی نام‌بُرده، تنظیم مجدد شده‌اند. سپس با عبور فریم‌های ویدئویی آموزشی از این شبکه، ماتریس‌های ویژگیِ لایه اولِ این شبکه به ازای هر فریم، به‌صورت برداری ذخیره شده‌اند. در نهایت با آموزش مدل رگرسیونی، میان داده‌های fMRI و ماتریس‌های ویژگی مربوط به هر تصویر، به مدلی که، تبدیل کننده‌یِ داده‌های fMRI به ماتریس‌های ویژگی تصاویر تست است، دست یافته‌اند. نمونه‌ای از نتایج بصری حاصل شده از این مطالعه در شکل ۵-۱۶ قابل مشاهده هستند.

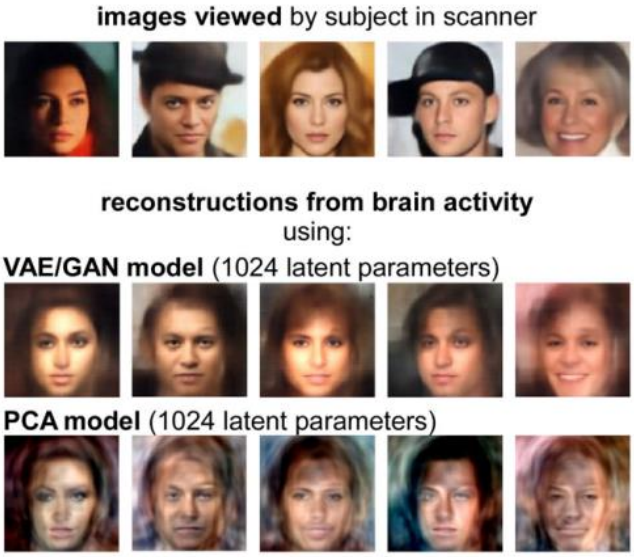


شکل ۵-۱۶: نتایج کیفی مدل رمزگشایی مغز پیشنهادی در اولین مطالعه مرجع [۹۲]. در قسمت بالا تصاویر اصلی نمایش داده شده است. در قسمت پایین بازسازی‌های به‌دست آمده قابل مشاهده هستند.

براساس تنها یکسان بودن مجموعه داده‌های استفاده شده در مطالعه [۹۲] و پژوهش ما، می‌توان قیاسی بر روش‌های پیشنهاد شده داشت. نتیجه‌ی این مقایسه بر عهده مخاطب و درک وی از تصاویر بازساری شده در خروجی هر کدام از مطالعه‌ها است.

دومین مطالعه [۹۳] الهام‌بخش در زمینه‌ی ایده به‌کار گرفته شده در پژوهش ما بوده است. به این معنا که در آن نیز، از نگاشت داده‌های fMRI بر کدهای نهان تصاویر به عنوان ایده اصلی یاد

شده است. با این تفاوت که در گام اول، جهت ایجاد کدهای نهان، از ترکیب دو شبکه VAE و GAN استفاده شده است. و در گام دوم نیز از دو رویکرد بر پایه رگرسیون خطی و همچنین از آنالیز مولفه اصلی[۱] (PCA) بهره گرفته‌اند. این مطالعه بر روی مجموعه داده‌های fMRI با مشاهده تصاویر چهره انجام شده است. به علت تفاوت در الگوریتم‌ها و همچنین مشخصات آموزشی متفاوتِ این مطالعه و پژوهش ما، مقایسه نتایج، چندان امر قابل قبولی نخواهد بود. با این وجود، نمونه‌هایی از تصاویر به‌دست آمده در خروجی این مطالعه را در شکل ۵-۱۷ می‌توان مشاهده کرد.



شکل ۵-۱۷: نتایج کیفی مدل رمزگشایی مغز پیشنهادی در دومین مطالعه مرجع [۹۳]. در قسمت بالایی تصاویر اصلی نمایش داده شده است. در قسمت میانی تصاویر بازسازی شده توسط مدل نگاشت مبتنی‌بر رگرسیون خطی قابل مشاهده است. قسمت پایینی تصاویر بازسازی شده توسط مدل نگاشت مبتنی بر آنالیز مو لفه اصلی را نمایش می‌دهد.

[۱] Principal Component Analysis

۶- فصل ششم

نتیجه گیری و پیشنهادات

۶-۱- مقدمه

در فصل پیشین نتایج به‌دست آمده از رویکرد پیشنهادی، ارائه و به تفصیل تحلیل شدند. در این فصل برآنیم مطالب پایان نامه را جمع بندی و نتیجه‌گیری کنیم، طوری‌که نقاط قوت و ضعف این روش شناخته شود و پیشنهاداتی برای بهبود این رویکرد ارائه کنیم.

۶-۲- نتیجه‌گیری

در این پژوهش هدف ما ارائه رویکردی بود که به کمک آن بتوانیم آن‌چه که انسان می‌بیند و درک می‌کند را از روی داده‌های fMRI، بازسازی کنیم. هدفی که به "خواندن ذهن[۱]" نیز موسوم است.

روش پیشنهاد شده در این پژوهش، مبتنی‌بر شبکه‌های عصبی عمیق بوده و در دو گام متوالی آموزش داده می‌شود. در گام اول، با به‌کارگیری شبکه‌های خودرمزگذار رقابتی، به فضای نهان تصاویر دست یافته و در گام دوم نیز با دو روش رگرسیون خطی و شبکه‌های عصبی، داده‌های fMRI، بر فضای نهان تصاویر نگاشت می‌شوند. در نهایت، مدل آموزش دیده قادر است تصاویر مشاهده شده توسط سوژه را با انتقال داده‌های fMRI به فضای نهان تصاویر، بازسازی کند.

رویکرد پیشنهادی قادر به بازسازی تصاویر در ۱۵ گروه مختلف است. این گروه ها شامل محیط سرپوشیده، محیط باز، انسان، چهره، پرنده، حشره، حیوانات آب‌زی، حیوانات خشکی‌زی، گل، میوه، چشم‌اندازهای طبیعی، ماشین، هواپیما، کشتی و ورزش کردن هستند. گروه‌های انتخابی در این پژوهش، از جمله مواردی هستند که هر انسانی روزانه و به‌طور معمول با آن‌ها مواجه می‌شود. به‌علاوه در قالب ویدئو در اختیار سوژه‌های آزمایشی قرار گرفته‌اند. پویایی و اطلاعات بالای این تصاویر بر دشواری رمزگشایی مغز می‌افزاید. این درحالی است که، رویکرد پیشنهادی برای این مسئله عملکرد قابل توجهی داشته و بازنمایی‌های قابل درکی را ارائه داده است. با توجه به نتایج به دست

[۱] Mind reading

آمده، می‌توان دریافت که شبکه‌های عصبی عمیق از قدرت خوبی جهت فراهم کردن فضای نهان تصاویر، به‌منظور رمزگشایی مغز، برخوردار هستند. نتایج تجربی نشان داد که مدل ساخته شده براساس شبکه عصبی عملکرد بهتری نسبت به مدل ساخته شده براساس نگاشت مبتنی‌بر رگرسیون خطی داشت.

۶-۳- پیشنهادات

- بی‌شک رمزگشایی مغز، در موارد درون کلاسی امر دشواری خواهد بود. به عنوان زمینه تحقیقاتی آتی، می‌توان ایده‌ی این پایان نامه را با تغییراتی، روی تصاویر درون کلاسی مثلا تصاویر مختلف چهره، پیاده کرد.
- در این پژوهش، ساختار شبکه خودرمزگذار رقابتی به صورت تجربی انتخاب شده است. بررسی کامل‌تر روی ساختار شبکه و بهبود سرعت و دقت عملکرد می‌تواند موضوع دیگری برای کارهای آتی باشد.
- در رویکرد پیشنهادی، از جمله پیش‌پردازش‌های انجام گرفته بر روی تصاویر، انتقال آن‌ها به سطح خاکستری است. به عبارتی روش پیشنهادی با چنین فرضی پیاده شده است. این در حالی است که، در عمل، رنگ، در درک بینایی انسان بسیار موثر و تعیین کننده است. به طوری‌که سیستم بینایی انسان براساس تنوع رنگ‌ها ممکن است، درک متفاوتی از تصاویر با ساختارهای یکسان داشته باشد. بنابراین تمرکز بر روی تصاویر رنگی، می‌تواند از جمله پیشنهادات در جهت رشد این زمینه باشد.
- در اینجا تنها به رمزگشایی مغز پرداخته شده است. درحالی که ترکیب آن با رویکردی جهت رمزگذاری مغز، به نوعی قادر به مدلسازی سامانه بینایی انسان خواهد بود. به‌گونه‌ای که سیگنال‌های fMRI توسط مدل رمزگذار تولید شوند و در مدل رمزگشا به عنوان ورودی مورد استفاده قرار گیرند.

منابع

[۱] ع. امیری, (۱۳۹۰), پایان‌نامه ارشد: "بازشناسی اشیا مبتنی‌بر سیستم بینایی انسان," دانشکده مهندسی برق و کامپیوتر, دانشگاه تربیت دبیر شهید رجائی.

[۲] ر. ابراهیم‌پور, (۱۳۹۵), *بازشناسی اشیاء علوم اعصاب بینایی، مدل های محاسباتی-شناختی، یادگیری عمیق*, چاپ اول. انتشارات دانشگاه تربیت دبیر شهید رجائی.

[3] M. D. Levine, (1985), *Vision in man and machine*. McGraw-Hill College.

[4] S. H. C. Hendry, R. C. Reid, P. Pether, S. H. C. Hendry, and R. C. Reid, (2000), "The koniocellular pathway in primate vision," *Annu. Rev. Neurosci.*, vol. 23, no. 1, pp. 127–153.

[5] J. Kubilius, D. Linsley, and C. Hill, (2015), "Encoding Models for High-level Visual Features," Report.

[6] D. D. Cox and T. Dean, (2014), "Neural networks and neuroscience-inspired computer vision," *Curr. Biol.*, vol. 24, no. 18, pp. R921–R929.

[7] D. L. K. Yamins and J. J. DiCarlo, (2016), "Using goal-driven deep learning models to understand sensory cortex," *Nat. Neurosci.*, vol. 19, no. 3, pp. 356–365.

[8] N. Kriegeskorte, (2015), "Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing," *Annu. Rev. Vis. Sci.*, vol. 1, no. 1, pp. 417–446.

[9] Y. Lecun, Y. Bengio, and G. Hinton, (2015), "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444.

[10] J. Y. H. Ng *et al.*, (2015), "Beyond short snippets: Deep networks for video classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 07-12-June, pp. 4694–4702.

[11] A. Garcia-Perez *et al.*, (2015), "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

[12] H. Wen, J. Shi, Y. Zhang, K. H. Lu, J. Cao, and Z. Liu, (2018), "Neural encoding and decoding with deep learning for dynamic natural vision," *Cereb. Cortex*, vol. 28, no. 12, pp. 4136–4160.

[13] U. Güçlü and M. A. J. J. van Gerven, (2015), "Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream," *J. Neurosci.*, vol. 35, no. 27, pp. 10005–10014.

[14] D. L. K. Yamins, H. Hong, C. F. Cadieu, E. A. Solomon, D. Seibert, and J. J. DiCarlo, (2014), "Performance-optimized hierarchical models predict neural responses in higher visual cortex," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 23, pp. 8619–8624.

[15] S.-M. M. Khaligh-Razavi and N. Kriegeskorte, (201), "Deep supervised, but not unsupervised, models may explain IT cortical representation," *PLoS*

Comput. Biol., vol. 10, no. 11, p. e1003915.

[16] T. Naselaris, K. N. Kay, S. Nishimoto, and J. L. Gallant, (2011), "Encoding and decoding in fMRI," *Neuroimage*, vol. 56, no. 2, pp. 400–410.

[17] D. P. Kingma and M. Welling, (2013), "Auto-encoding variational bayes," *arXiv Prepr. arXiv1312.6114*, no. Ml, pp. 1–14.

[18] D. J. Rezende *et al.*, (2014), "Stochastic backpropagation and approximate inference in deep generative models," *arXiv*, vol. 4, p. arXiv-1401.

[19] I. J. Goodfellow *et al.*, (2014), "Generative adversarial nets," in *Advances in neural information processing systems*, vol. 3, no. January, pp. 2672–2680.

[20] Y. Güçlütürk, U. Güçlü, K. Seeliger, S. Bosch, R. Van Lier, and M. Van Gerven, (2017), "Reconstructing perceived faces from brain activations with deep adversarial neural decoding," *Adv. Neural Inf. Process. Syst.*, vol. 2017-December, no. Nips, pp. 4247–4258.

[21] C. C. Du, C. C. Du, and H. He, (2017), "Sharing deep generative representation for perceived image reconstruction from human brain activity," in *2017 International Joint Conference on Neural Networks (IJCNN)*, vol. 2017-May, pp. 1049–1056.

[22] K. Han, H. Wen, J. Shi, K.-H. Lu, Y. Zhang, and Z. Liu, (2018), "Variational autoencoder: An unsupervised model for modeling and decoding fMRI activity in visual cortex," *bioRxiv*, no. November, p. 214247.

[23] K. Seeliger, U. Güçlü, L. Ambrogioni, Y. Güçlütürk, and M. A. J. J. van Gerven, (2018), "Generative adversarial networks for reconstructing natural images from brain activity," *Neuroimage*, vol. 181, pp. 775–785.

[24] G. St-Yves and T. Naselaris, (2018), "Generative adversarial networks conditioned on brain activity reconstruct seen images," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1054–1061.

[25] G. Shen, K. Dwivedi, K. Majima, T. Horikawa, and Y. Kamitani, (2019), "End-to-end deep image reconstruction from human brain activity," *Front. Comput. Neurosci.*, vol. 13, no. April, p. 21.

[26] Y. Kamitani and F. Tong, (2005), "Decoding the visual and subjective contents of the human brain," *Nat. Neurosci.*, vol. 8, no. 5, pp. 679–685.

[27] J. D. Haynes and G. Rees, (2006), "Decoding mental states from brain activity in humans," *Nat. Rev. Neurosci.*, vol. 7, no. 7, pp. 523–534.

[28] K. He, X. Zhang, S. Ren, and J. Sun, (2016), "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, vol. 2016-Decem, pp. 770–778.

[29] B. B. Averbeck, P. E. Latham, and A. Pouget, (2006), "Neural correlations, population coding and computation," *Nat. Rev. Neurosci.*, vol. 7, no. 5, pp. 358–366.

[۳۰] س. م. م. صالحی, (۱۳۹۰), پایان‌نامه ارشد: "تفسیر تصاویر حاصل از fMRI به منظور تحلیل

حالات و رفتار انسانی," دانشکده مهندسی کامپیوتر و فناوری اطلاعات, دانشگاه صنعتی شاهرود.

[31] "Modeling fMRI." http://gureckislab.org/courses/fall19/labincp/labs/lab1mri-pt1.html.

[32] S. Ogawa and T. -M Lee, (1990), "Magnetic resonance imaging of blood vessels at high fields: In vivo and in vitro measurements and image simulation," *Magn. Reson. Med.*, vol. 16, no. 1, pp. 9–18.

[33] N. K. Logothetis and J. Pfeuffer, (2004), "On the nature of the BOLD fMRI contrast mechanism," *Magn. Reson. Imaging*, vol. 22, no. 10 SPEC. ISS., pp. 1517–1531.

[34] "Magnetism - Questions and Answers in MRI." http://mriquestions.com/does-boldbrain-activity.html.

[35] P. Bogorodzki, J. Rogowska, and D. A. Yurgelun-Todd, (2005), "Structural group classification technique based on regional fMRI BOLD responses," *IEEE Trans. Med. Imaging*, vol. 24, no. 2, pp. 389–398.

[36] M. Chiew and S. J. Graham, (2011), "BOLD contrast and noise characteristics of densely sampled multi-echo fMRI data," *IEEE Trans. Med. Imaging*, vol. 30, no. 9, pp. 1691–1703.

[37] Z. Chen and V. Calhoun, (2011), "A computational multiresolution BOLD fMRI model," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 10 PART 2, pp. 2995–2999.

[38] D. J. Jacobsen, K. H. Madsen, and L. K. Hansen, (2006), "Identification of non-linear models of neural activity in bold fmri," in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pp. 952–955.

[39] T. Jin and S.-G. G. Kim, (2006), "Spatial dependence of CBV-fMRI: a comparison between VASO and contrast agent based methods," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, no. 10, pp. 25–28.

[40] J. Deng *et al.*, (2009), "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, no. May 2014, pp. 248–255.

[41] G. E. Hinton and T. J. Sejnowski, (1986), "Learning and relearning in Boltzmann machines," *Parallel Distrib. Process. Explor. Microstruct. Cogn.*, vol. 1, no. 282–317, p. 2.

[42] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, (2015), "Adversarial Autoencoders," *arXiv Prepr. arXiv1511.05644*.

[43] C. Du, J. Li, L. Huang, and H. He, (2019), "Brain Encoding and Decoding in fMRI with Bidirectional Deep Generative Models," *Engineering*, vol. 5, no. 5, pp. 948–953.

[44] K. N. Kay, J. Winawer, A. Rokem, A. Mezer, and B. A. Wandell, (2013), "A two-stage cascade model of BOLD responses in human visual cortex," *PLoS*

Comput Biol, vol. 9, no. 5, p. e1003079.

[45] G. St-Yves and T. Naselaris, (2018), "The feature-weighted receptive field: an interpretable encoding model for complex feature spaces," *Neuroimage*, vol. 180, pp. 188–202.

[46] S. Schoenmakers, U. Güçlü, M. van Gerven, and T. Heskes, (2015), "Gaussian mixture models and semantic gating improve reconstructions from human brain activity," *Front. Comput. Neurosci.*, vol. 8, no. JAN, pp. 1–10.

[47] A. S. Cowen, M. M. Chun, and B. A. Kuhl, (2014), "Neural portraits of perception: reconstructing face images from evoked brain activity," *Neuroimage*, vol. 94, pp. 12–22.

[48] H. Lee and B. A. Kuhl, (2016), "Reconstructing perceived and retrieved faces from activity patterns in lateral parietal cortex," *J. Neurosci.*, vol. 36, no. 22, pp. 6069–6082.

[49] T. Horikawa and Y. Kamitani, (2017), "Hierarchical neural representation of dreamed objects revealed by brain decoding with deep neural network features," *Front. Comput. Neurosci.*, vol. 11, no. January, pp. 1–11.

[50] T. Naselaris, C. A. Olman, D. E. Stansbury, K. Ugurbil, and J. L. Gallant, (2015), "A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes," *Neuroimage*, vol. 105, pp. 215–228.

[51] P. Zeidman, E. H. Silson, D. S. Schwarzkopf, C. I. Baker, and W. Penny, (2018), "Bayesian population receptive field modelling," *Neuroimage*, vol. 180, pp. 173–187.

[52] J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten, and P. Pietrini, (2001), "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science (80-.).*, vol. 293, no. 5539, pp. 2425–2430.

[53] T. Naselaris, R. J. Prenger, K. N. Kay, M. Oliver, and J. L. Gallant, (2009), "Bayesian Reconstruction of Natural Images from Human Brain Activity," *Neuron*, vol. 63, no. 6, pp. 902–915.

[54] T. Horikawa, M. Tamaki, Y. Miyawaki, and Y. Kamitani, (2013), "Neural decoding of visual imagery during sleep," *Science (80-.).*, vol. 340, no. 6132, pp. 639–642.

[55] Y. Miyawaki *et al.*, (2008), "Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders," *Neuron*, vol. 60, no. 5, pp. 915–929.

[56] Y. Fujiwara, Y. Miyawaki, and Y. Kamitani, (2013), "Modular encoding and decoding models derived from Bayesian canonical correlation analysis," *Neural Comput.*, vol. 25, no. 4, pp. 979–1005.

[57] S. Yu, N. Zheng, Y. Ma, H. Wu, and B. Chen, (2019), "A Novel Brain Decoding Method: A Correlation Network Framework for Revealing Brain Connections," *IEEE Trans. Cogn. Dev. Syst.*, vol. 11, no. 1, pp. 95–106.

[58] S. Schoenmakers, M. Barth, T. Heskes, and M. van Gerven, (2013), "Linear reconstruction of perceived images from human brain activity," *Neuroimage*,

vol. 83, pp. 951–961.

[59] K. N. Kay, T. Naselaris, R. J. Prenger, and J. L. Gallant, (2008), "Identifying natural images from human brain activity," *Nature*, vol. 452, no. 7185, pp. 352–355.

[60] T. Horikawa and Y. Kamitani, (2017), "Generic decoding of seen and imagined objects using hierarchical visual features," *Nat. Commun.*, vol. 8, no. May, pp. 1–15.

[61] C. Zhang *et al.*, (2018), "Constraint-free natural image reconstruction from fMRI signals based on convolutional neural network," pp. 1–16.

[62] S. Schoenmakers *et al.*, (2014), "Gaussian mixture models improve fMRI-based image reconstruction," in *2014 International Workshop on Pattern Recognition in Neuroimaging*, pp. 1–4.

[63] B. A. Krizhevsky, I. Sutskever, G. E. Hinton, A. Krizhevsky, I. Sutskever, and G. E. Hinton, (2017), "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90.

[64] E. L. Denton, S. Chintala, A. Szlam, and R. Fergus, (2015), "Deep generative image models using a laplacian pyramid of adversarial networks," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 1486–1494.

[65] L. Deng *et al.*, (2013), "Recent advances in deep learning for speech research at Microsoft," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8604–8608.

[66] Q. Le and T. Mikolov, (2014), "Distributed representations of sentences and documents," *31st Int. Conf. Mach. Learn. ICML 2014*, vol. 4, pp. 2931–2939.

[67] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, (2012), "Decoding subject-driven cognitive states with whole-brain connectivity patterns," *Cereb. cortex*, vol. 22, no. 1, pp. 158–165.

[68] F. Mokhtari and G. Hossein-zadeh, (2013), "Decoding brain states using backward edge elimination and graph kernels in fMRI connectivity networks," *J. Neurosci. Methods*, vol. 212, no. 2, pp. 259–268.

[69] E. Yargholi and G.-A. Hossein-Zadeh, (2016), "Brain decoding-classification of hand written digits from fMRI data employing bayesian networks," *Front. Hum. Neurosci.*, vol. 10, no. July, p. 351.

[70] E. Yargholi and G.-A. Hossein-zadeh, (2016), "Reconstruction of digit images from human brain fMRI activity through connectivity informed Bayesian networks," *J. Neurosci. Methods*, vol. 257, pp. 159–167.

[71] J. R. Manning *et al.*, (2018), "A probabilistic approach to discovering dynamic full-brain functional connectivity patterns," *Neuroimage*, vol. 180, pp. 243–252.

[72] M. Chen, J. Han, X. Hu, X. Jiang, L. Guo, and T. Liu, (2014), "Survey of encoding and decoding of visual stimulus via FMRI: an image analysis perspective," *Brain Imaging Behav.*, vol. 8, no. 1, pp. 7–23.

[73] M. A. J. van Gerven, M. A. J. Van Gerven, and M. A. J. van Gerven, (2017), "A primer on encoding models in sensory neuroscience," *J. Math. Psychol.*, vol. 76, pp. 172–183.

[74] P.-C. Kuo, Y.-S. Chen, L.-F. Chen, and J.-C. Hsieh, (2014), "Decoding and encoding of visual patterns using magnetoencephalographic data represented in manifolds," *Neuroimage*, vol. 102, pp. 435–450.

[75] J. Schmidhuber, (2015), "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117.

[76] R. Matsumura, K. Harada, Y. Domae, W. Wan, W. S. McCulloch, and W. Pitts, (1943), "A logical calculus of the ideas immanent in nervous activity," *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133.

[77] R. M. Cichy, A. Khosla, D. Pantazis, A. Torralba, and A. Oliva, (2016), "Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence," *Sci. Rep.*, vol. 6, no. May, pp. 1–13.

[78] M. Eickenberg, A. Gramfort, G. Varoquaux, and B. Thirion, (2017), "Seeing it all: Convolutional network layers map the function of the human visual system," *Neuroimage*, vol. 152, pp. 184–194.

[79] J. J. DiCarlo, D. Zoccolan, and N. C. Rust, (2012), "How does the brain solve visual object recognition?," *Neuron*, vol. 73, no. 3, pp. 415–434.

[80] J. J. DiCarlo and D. D. Cox, (2007), "Untangling invariant object recognition," *Trends Cogn. Sci.*, vol. 11, no. 8, pp. 333–341.

[81] J. Li, Z. Zhang, and H. He, (2016), "Visual information processing mechanism revealed by fMRI data," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9919 LNAI, pp. 85–93.

[82] I. Higgins *et al.*, (2016), "Early visual concept learning with unsupervised deep learning," *arXiv Prepr. arXiv1606.05579*.

[83] T. D. Kulkarni, W. F. Whitney, P. Kohli, and J. B. Tenenbaum, (2015), "Deep convolutional inverse graphics network," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 2539–2547.

[84] S. M. Ali Eslami *et al.*, (2016), "Attend, infer, repeat: Fast scene understanding with generative models," in *Advances in Neural Information Processing Systems*, pp. 3225–3233.

[85] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, (2006), "Beyond mind-reading: multi-voxel pattern analysis of fMRI data," *Trends Cogn. Sci.*, vol. 10, no. 9, pp. 424–430.

[86] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, (2017), "Image-to-image translation with conditional adversarial networks," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5967–5976.

[87] M. Y. Liu, T. Breuel, and J. Kautz, (2017), "Unsupervised image-to-image translation networks," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no.

Nips, pp. 701–709.

[88] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, (2016), "Generative adversarial text to image synthesis," *33rd Int. Conf. Mach. Learn. ICML 2016*, vol. 3, pp. 1681–1690.

[89] S. Hong, D. Yang, J. Choi, and H. Lee, (2018), "Inferring Semantic Layout for Hierarchical Text-to-Image Synthesis," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, no. Figure 1, pp. 7986–7994.

[90] D. He *et al.*, (2016), "Dual learning for machine translation," *Adv. Neural Inf. Process. Syst.*, no. Nips, pp. 820–828.

[91] Y. Xia, T. Qin, W. Chen, J. Bian, N. Yu, and T. Y. Liu, (2017), "Dual supervised learning," *34th Int. Conf. Mach. Learn. ICML 2017*, vol. 8, no. 1, pp. 5792–5803.

[92] H. Wen, J. Shi, Y. Zhang, K. H. Lu, J. Cao, and Z. Liu, (2018), "Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision," *Cereb. Cortex*, vol. 28, no. 12, pp. 4136–4160.

[93] R. VanRullen and L. Reddy, (2019), "Reconstructing faces from fMRI patterns using deep generative neural networks," *Commun. Biol.*, vol. 2, no. 1, pp. 1–10.

Abstract

In recent decades, research on encoding and decoding brain activity from functional magnetic resonance imaging (fMRI) has made significant progress. However, decoding brain activity in order to reconstruct natural images perceived by humans remains an issue to be investigated, as these images contain various and complex information.

In this study, an approach is proposed to decode brain activity based on Adversarial Auto-Encoder networks (AAEs). Hence, at first, an AAE neural network is trained using more than 11k natural images not seen by the subjects. The AAE latent space provides a meaningful description of each image. We then developed a mapping between fMRI patterns and latent space, according to which the fMRI patterns are transformed to latent code in the same dimension as the AAE codes. When testing the decoding model, we use trained mapping to transform the fMRI patterns of all three human subjects into latent codes. Finally, the codes are transformed into intelligible representations of the test images using the decoding structure of the AAE network. The proposed decoding model is trained and tested using separate data.

In order to quantify the performance of the proposed method, multi-scale structural similarity and correlation coefficient are reported for the reconstructed images. The obtained results indicate the successful and promising performance of the proposed method. So that for two subjects, the average correlation coefficient on the test images was reported to be more than 0.7.

Keywords: brain encoding and decoding, functional Magnetic Resonance Imaging, Adversarial Auto-Encoder networks, Multi-scale structural similarity, Correlation Coefficient.

Shahrood University of Technology

Faculty of Electrical Engineering

MSc Thesis in: Electrical Engineering-Telecommunication

Reconstruction of Perceived Images from Functional Magnetic Resonance Imaging Based on Neural Networks

By: Saeedeh poorghasemian

Supervisor

Dr. Hossein Khosravi

December 2020