



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گزارش پروژه کارشناسی

گرایش هوش مصنوعی

تشخیص بیماری‌های ریوی از روی تصاویر پزشکی با
استفاده از یادگیری خودنظراتی

نگارش

میتراعمرانی

استاد راهنما

دکتر پیمان ادبی

شهریور ۱۴۰۳

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشگاه اصفهان

دانشکده مهندسی کامپیوتر

گروه هوش مصنوعی

پروژه کارشناسی رشته‌ی مهندسی کامپیوتر گرایش هوش مصنوعی

..... آقای / خانم

تحت عنوان

.....

..... در تاریخ / /

توسط هیأت داوران زیر بررسی و با نمره به تصویب نهایی رسید.

استاد راهنمای پژوهش:

دکتر

..... امضا:

استاد داور:

دکتر

..... امضا:

امضای مدیر گروه

سپاس گزاری

سپاس گزارم از زحمات پدر، مادر، مینای عزیزم و استاد محترم که با راهنمایی های ارزشمند شان، همواره کنار من بودند و همه عزیزانی که با حضور و
یاریشان من را در انجام این پروژه همراهی کردند.

یتراعمرانی

۱۴۰۳ شهریور

چکیده

در این پروژه، ابتدا مدل‌های یادگیری نظارت شده با استفاده از مجموعه داده‌های برچسب‌گذاری شده مربوط به تصاویر ریه پیاده‌سازی شده‌اند که در این مرحله، مدل‌های متداول شبکه‌های عصبی عمیق مانند VGG16 و ResNet50 بر روی این تصاویر آموزش داده شده و عملکرد آن‌ها در تشخیص بیماری‌های ریه ارزیابی شده است. سپس، روش‌های یادگیری خودناظارتی نیز پیاده‌سازی شده‌اند تا به مدل‌ها امکان دهنده با استفاده از تصاویر بدون برچسب و یادگیری ویژگی‌های مفید، کارایی خود را در شرایط کمبود داده‌های برچسب‌گذاری شده بهبود بخشد. برای این منظور، از مجموعه داده‌ای شامل تصاویر رادیوگرافی و سی‌تی‌اسکن مربوط به بیماری‌های ریه استفاده شده و پس از پیاده‌سازی هر دو رویکرد، مقایسه‌ای بین عملکرد مدل‌های یادگیری با نظارت و یادگیری خودناظارتی انجام شده تا تفاوت‌ها و مزایای هر روش در شرایط مختلف مورد بررسی قرار گیرد.

واژه‌های کلیدی:

تصاویر پزشکی، شبکه‌های عصبی عمیق، یادگیری خودناظارتی، تشخیص بیماری‌های ریوی، یادگیری انتقالی، بینایی ماشین.

فهرست مطالب

صفحه

عنوان

۱	پیش گفتار	۱
۲	۱-۱ مقدمه	
۳	۲-۱ بیان مسئله	
۴	۳-۱ معرفی پروژه	
۴	۴-۱ ساختار پایان نامه	
۵	۲ مفاهیم	۲
۶	۱-۲ شبکه‌های عصبی پیچشی	
۶	۱-۱-۲ معرفی شبکه‌های عصبی پیچشی	
۶	۲-۱-۲ اجزای اصلی شبکه‌های عصبی پیچشی	
۸	۳-۱-۲ آموزش شبکه‌های عصبی پیچشی	
۸	۴-۱-۲ انواع شبکه‌های عصبی پیچشی	
۸	۲-۲ یادگیری انتقالی	
۹	۱-۲-۲ ایده کلی یادگیری انتقالی	
۱۱	۲-۲-۲ مزایای یادگیری انتقالی	
۱۱	۳-۲-۲ چالش‌های موجود در نبود داده‌های کافی	
۱۱	۴-۲-۲ استفاده از مدل‌های پیش‌آموزش‌دیده	
۱۱	۳-۲ یادگیری خودنظرارتی	
۱۱	۱-۳-۲ تعریف یادگیری خودنظرارتی	
۱۲	۲-۳-۲ ریشه یادگیری خودنظرارتی	
۱۴	۳-۳-۲ تکامل خانواده‌های یادگیری خودنظرارتی	
۱۶	۳ شرح پروژه	
۱۷	۱-۳ جمع آوری داده	
۱۷	۱-۱-۳ معرفی مجموعه داده‌ها	
۱۸	۲-۳ پیش‌پردازش داده‌ها و ایجاد مجموعه داده	
۱۸	۱-۲-۳ تغییر نام و ساختاردهی مجموعه داده	
۱۸	۲-۲-۳ تقسیم مجموعه داده	
۱۹	۳-۲-۳ ترکیب مجموعه داده‌ها	
۲۰	۴-۲-۳ رفع عدم تعادل دسته‌ها	
۲۱	۵-۲-۳ نمای کلی از مجموعه داده نهایی	
۲۲	۳-۳ یادگیری نظارت شده	
۲۲	۱-۳-۳ مرحله ۱: انتخاب مدل و یادگیری انتقالی	

۲۳	۲-۳-۳ مرحله ۲: تنظیم دقیق مدل
۲۴	۳-۳-۳ مرحله ۳: ارزیابی مدل و توقف زودهنگام
۲۴	۴-۳-۳ مرحله ۴: جمع‌آوری معیارها و نظارت بر عملکرد
۲۴	۵-۳-۳ مرحله ۵: تست مدل
۲۴	۴-۳ یادگیری خودناظارتی
۲۵	SimCLR ۱-۴-۳
۲۹	Barlow Twins ۲-۴-۳
۳۲	۴ نتایج
۳۳	۱-۴ مقدمه
۳۳	۲-۴ مدل‌های یادگیری با نظارت
۳۳	۱-۲-۴ نتایج مدل ResNet50
۳۷	۲-۲-۴ نتایج مدل VGG16
۴۰	۳-۴ یادگیری خودناظارتی
۴۲	۱-۳-۴ نتایج مدل SimCLR با رمزگذار آزاد
۴۵	۲-۳-۴ نتایج مدل SimCLR با رمزگذار ثابت
۴۸	۳-۳-۴ نتایج مدل Barlow Twins با رمزگذار آزاد
۵۰	۴-۳-۴ نتایج مدل Barlow Twins با رمزگذار ثابت
۵۳	۵ جمع‌بندی و نتیجه‌گیری
۵۴	۱-۵ مدل‌های بررسی شده
۵۴	۲-۵ تحلیل عملکرد هر مدل
۵۴	۱-۲-۵ مدل‌های یادگیری نظارت شده
۵۵	۲-۲-۵ مدل‌های SimCLR
۵۵	۳-۲-۵ مدل‌های Barlow Twins
۵۶	۳-۵ مقایسه دقت مدل‌ها با یکدیگر
۵۶	۴-۵ مقایسه زمانی مدل‌ها با یکدیگر
۵۷	۵-۵ مقایسه مدل‌های یادگیری نظارت شده با مدل‌های یادگیری خودناظارتی
۵۷	۱-۵-۵ دقت و عملکرد
۵۷	۲-۵-۵ ثبات آموزش و کاهش خطای
۵۷	۳-۵-۵ زمان آموزش و هزینه محاسباتی
۵۸	۴-۵-۵ هزینه برچسب‌گذاری و کاربرد پذیری
۵۸	۶-۵ نتیجه‌گیری نهایی: ارزش مدل‌های خودناظارتی
۵۹	منابع و مراجع

فهرست اشکال

صفحه

شکل

۶	۱-۲ شبکه عصبی پیچشی [۱]
۹	۲-۲ یادگیری انتقالی [۲]
۱۲	۳-۲ مراحل روش یادگیری خودناظارتی [۳]
۱۹	۲-۳ مجموعه داده بعد از پیش پردازش
۱۹	۱-۳ مجموعه داده قبل از پیش پردازش
۲۱	۳-۳ دسته‌ها قبل از متعادل‌سازی
۲۱	۴-۳ دسته‌ها بعد از متعادل‌سازی
۲۲	۵-۳ معماری شبکه ResNet50 [۴]
۲۳	۶-۳ معماری شبکه VGG16 [۵]
۲۵	۷-۳ افزایش‌های تصویر در SimCLR [۶]
۲۷	۸-۳ چارچوب ساده برای یادگیری تمایزی بازنمایی‌های بصری [۷]
۲۸	۹-۳ معماری SimCLR [۸]
۳۱	۱۰-۳ معماری Barlow Twins [۹]
۳۴	۱-۴ خطای آموزش، اعتبارسنجی و آزمایش ResNet50
۳۴	۲-۴ دقیق آموزش و اعتبارسنجی و آزمایش ResNet50
۳۵	۳-۴ صحیح آموزش و اعتبارسنجی و آزمایش ResNet50
۳۵	۴-۴ یادآوری آموزش و اعتبارسنجی و آزمایش ResNet50
۳۶	۵-۴ امتیاز F1 آموزش و اعتبارسنجی و آزمایش ResNet50
۳۷	۶-۴ خطای آموزش، اعتبارسنجی و آزمایش VGG16
۳۷	۷-۴ دقیق آموزش و اعتبارسنجی و آزمایش VGG16
۳۸	۸-۴ صحیح آموزش و اعتبارسنجی و آزمایش VGG16
۳۸	۹-۴ یادآوری آموزش و اعتبارسنجی و آزمایش VGG16
۳۹	۱۰-۴ امتیاز F1 آموزش و اعتبارسنجی و آزمایش VGG16
۴۰	۱۱-۴ خطای آموزش در طول دوره‌ها در SimCLR
۴۱	۱۲-۴ خطای آموزش در طول دوره‌ها در Barlow Twins
۴۲	۱۳-۴ خطای آموزش در طول دوره‌ها در SimCLR آزاد
۴۳	۱۴-۴ دقیق آموزش در طول دوره‌ها مدل SimCLR آزاد
۴۳	۱۵-۴ معیارهای ارزیابی آموزش در دوره نهایی در SimCLR آزاد
۴۵	۱۶-۴ خطای آموزش در طول دوره‌ها در SimCLR ثابت
۴۶	۱۷-۴ دقیق آموزش در طول دوره‌ها مدل SimCLR ثابت
۴۶	۱۸-۴ معیارهای ارزیابی آموزش در دوره نهایی در SimCLR ثابت

۴۸	۱۹-۴ خطای آموزش در طول دوره‌ها در Barlow Twins آزاد
۴۹	۲۰-۴ دقت آموزش در طول دوره‌ها مدل Barlow Twins آزاد
۴۹	۲۱-۴ معیارهای ارزیابی آموزش در دوره نهایی در Barlow Twins آزاد
۵۰	۲۲-۴ خطای آموزش در طول دوره‌ها در Barlow Twins ثابت
۵۱	۲۳-۴ دقت آموزش در طول دوره‌ها مدل Barlow Twins ثابت
۵۱	۲۴-۴ معیارهای ارزیابی آموزش در دوره نهایی در Barlow Twins ثابت

صفحه

فهرست جداول

جدول

- | | |
|---|----|
| ۱-۵ مقایسه دقیق نهایی مدل‌ها | ۵۶ |
| ۲-۵ مقایسه زمان‌های آموزش و استنتاج مدل‌ها به همراه توقف زودهنگام و تعداد دورها | ۵۶ |

فهرست نمادها

نماد مفهوم فضای اقلیدسی با بعد n

فصل اول

پیش گفتار

۱-۱ مقدمه

پردازش تصویر^۱، به عنوان شاخه‌ای از هوش مصنوعی^۲ و بینایی کامپیوتر^۳، نقش حیاتی در دنیای مدرن ایفا می‌کند. این علم کاربردهای گسترده‌ای دارد و یکی از زیرمجموعه‌های مهم آن، پردازش تصاویر پزشکی است که به‌ویژه برای تحلیل تصاویر مرتبط با آناتومی و عملکرد بدن انسان استفاده می‌شود. تصاویر پزشکی، شامل رادیوگرافی‌ها، سی‌تی اسکن‌ها، ام‌آرآی‌ها و تصاویر سونوگرافی، اطلاعات ارزشمندی را برای تشخیص بیماری‌ها، نظارت بر پیشرفت درمان و برنامه‌ریزی جراحی‌ها فراهم می‌کند.

یکی از مهم‌ترین انواع تصویربرداری پزشکی، تصاویر رادیوگرافی قفسه سینه (CXR)^۴ است که به طور گسترده برای تشخیص انواع بیماری‌های تنفسی از جمله سل (TB)^۵ و ذات‌الریه^۶ استفاده می‌شود. سل، یک بیماری عفونی بالقوه شدید، عمدتاً ریه‌ها را تحت تأثیر قرار می‌دهد و یکی از علل اصلی مرگ و میر در سراسر جهان، به ویژه در کشورهای کم‌درآمد است. ذات‌الریه، یک عفونت که باعث التهاب کیسه‌های هوایی در یکی یا هر دو ریه می‌شود، نیز یک عامل مهم بیماری و مرگ و میر در سطح جهان است. تشخیص دقیق این شرایط برای درمان و مدیریت مؤثر حیاتی است، اما تفسیر تصاویر رادیوگرافی قفسه سینه چالش‌برانگیز است و نیاز به تخصص بالایی دارد. یادگیری ماشین^۷، به ویژه شبکه‌های عصبی عمیق^۸ مانند ResNet50 [۱۰] و VGG16 [۱۱]، در خود کارسازی تفسیر تصاویر رادیوگرافی قفسه سینه نتایج قابل توجهی نشان داده‌اند. این مدل‌ها که بر روی مجموعه داده‌های بزرگ آموزش دیده‌اند، می‌توانند با شناسایی الگوهای در تصاویر، بیماری‌ها را تشخیص داده و دسته‌بندی کنند. با این حال، موفقیت این مدل‌ها در کاربردهای بالینی به شدت به در دسترس بودن مجموعه داده‌های بزرگ و با کیفیت برچسب‌گذاری شده بستگی دارد. برچسب‌گذاری تصاویر پزشکی فرآیندی پرزمخت است که نیاز به داشت تخصصی دارد و این امر آن را به یک فعالیت پرهزینه و زمان بر تبدیل می‌کند. این محدودیتها موجب شده تا بررسی الگوهای یادگیری جایگزینی که می‌توانند واپسی به داده‌های برچسب‌گذاری شده را کاهش دهند و در عین حال عملکرد مدل را حفظ یا حتی بهبود بخشنده، ضروری شود.

¹Image Processing

²Artificial Intelligence

³Computer Vision

⁴Chest X-Ray

⁵Tuberculosis

⁶Pneumonia

⁷Machine learning

⁸Deep neural networks

۲-۱ بیان مسئله

مسئله‌ای که این پژوهه به دنبال حل آن است، چندوجهی بوده و شامل چالش‌های پزشکی و فنی مرتبط با تشخیص بیماری‌های تنفسی مانند سل و ذات‌الریه با استفاده از تصاویر رادیوگرافی قفسه سینه می‌باشد. سل و ذات‌الریه از مهم‌ترین چالش‌های بهداشت عمومی در سطح جهانی هستند.

با این حال، تشخیص این بیماری‌ها بر اساس تصاویر رادیوگرافی قفسه سینه آسان نیست. نشانه‌های بصری سل و ذات‌الریه در این تصاویر می‌توانند ظرفی باشند و حتی توسط رادیولوژیست‌های با تجربه به راحتی با دیگر شرایط اشتباه گرفته شوند. علاوه بر این، در بسیاری از مناطق که سل شایع است، کمبود رادیولوژیست‌های آموزش‌دهنده وجود دارد که منجر به تأخیر در تشخیص و درمان می‌شود. سیستم‌های خودکار مبتنی بر یادگیری ماشین می‌توانند یک راه حل بالقوه ارائه دهند، زیرا تفسیرهای ثابت و دقیقی از این تصاویر ارائه می‌دهند و از این رو می‌توانند از تصمیم‌گیری‌های بالینی، به ویژه در مناطق کممنابع، حمایت کنند.

با وجود این وعده، استقرار این مدل‌ها در محیط‌های بالینی واقعی با چالش‌های قابل توجهی مواجه است. یکی از مهم‌ترین چالش‌ها، نیاز به مجموعه داده‌های بزرگ و برچسب‌گذاری شده برای آموزش این مدل‌ها است. در تصویربرداری پزشکی، داده‌های برچسب‌گذاری شده اغلب کمیاب هستند زیرا برچسب‌گذاری آن‌ها نیاز به تخصص دارد که هزینه‌بر و زمان‌بر است. کمبود داده به ویژه در حوزه‌های خاص پزشکی که برخی بیماری‌ها نادر هستند، محدودیت بیشتری ایجاد می‌کند. همچنین، تنوع بالای تصاویر پزشکی، پیچیدگی آموزش مدل‌های یادگیری عمیق^۹ را افزایش می‌دهد.

برای مقابله با این چالش‌ها، راه حل‌هایی مانند یادگیری انتقالی^{۱۰} و یادگیری خودنظرارتی^{۱۱} پیشنهاد شده است. در یادگیری انتقالی، از یک مدل یادگیری عمیق که بر روی مجموعه داده‌ای بزرگ مانند ImageNet آموزش دیده است به عنوان نقطه شروع برای پردازش تصاویر پزشکی استفاده می‌شود و سپس مدل با داده‌های کوچکتر تنظیم دقیق می‌شود. از طرف دیگر، یادگیری خودنظرارتی یک رویکرد نوین است که به مدل‌ها اجازه می‌دهد تا از داده‌های بدون برچسب برای یادگیری ویژگی‌های مهم در تصاویر استفاده کنند. این امر می‌تواند نیاز به داده‌های برچسب‌گذاری شده را به طور قابل توجهی کاهش دهد و به مدل‌ها اجازه دهد تا از مجموعه داده‌های بزرگتر و متنوع‌تری برای یادگیری استفاده کنند.

⁹Deep Learning

¹⁰Transfer Learning

¹¹Self-Supervised Learning

۱-۳ معرفی پروژه

این پروژه بر روی مقابله با محدودیت‌های سنتی یادگیری نظارت شده^{۱۲} در زمینه تشخیص سل و ذات‌الریه با استفاده از تصاویر رادیوگرافی قفسه سینه متمرکز است. هدف اصلی این پروژه بررسی و مقایسه اثربخشی روش‌های یادگیری خودناظارتی با روش‌های یادگیری نظارت شده سنتی است. به طور خاص، پروژه به دنبال تعیین این است که آیا روش‌های یادگیری خودناظارتی می‌توانند عملکردی مشابه یا بهتر از یادگیری نظارت شده ارائه دهند، در حالی که به طور قابل توجهی وابستگی به داده‌های برچسب‌گذاری شده را کاهش می‌دهند.

این پروژه از دو مجموعه داده بزرگ و عمومی استفاده می‌کند؛ پایگاه داده رادیوگرافی قفسه سینه سل استفاده شده در [۱۲] و مجموعه داده تصاویر رادیوگرافی قفسه سینه (ذات‌الریه) استفاده شده در [۱۳]. این مجموعه داده‌ها منبع ارزشمندی برای آموزش و ارزیابی مدل‌های یادگیری ماشین فراهم می‌کنند. در ابتدا، این پروژه شامل آموزش دو مدل یادگیری عمیق، ResNet50 و VGG16، با استفاده از رویکرد سنتی یادگیری نظارت شده بود. این مدل‌ها بر روی مجموعه داده‌های برچسب‌گذاری شده آموزش داده شدند تا تصاویر رادیوگرافی قفسه سینه را به عنوان نمایانگر سل، ذات‌الریه یا حالت سالم طبقه‌بندی کنند. این مرحله به وابستگی مدل‌های یادگیری نظارت شده به حجم زیادی از داده‌های برچسب‌گذاری شده اشاره دارد که هم هزینه‌بر و هم زمان‌بر است.

برای رفع این محدودیت‌ها، پروژه همچنین به بررسی روش‌های یادگیری خودناظارتی می‌پردازد. به طور خاص، مدل‌هایی مانند SimCLR [۶] و Barlow Twins [۹] به صورت خودناظارتی آزمایش و آموزش داده شدند. در این رویکرد، مدل‌ها ابتدا بر روی کل مجموعه داده بدون تکیه بر برچسب‌ها پیش‌آموزش داده شدند و به آن‌ها اجازه داده شد تا الگوهای ذاتی موجود در تصاویر را یاد بگیرند و رمزگذاری کنند. پس از این پیش‌آموزش، مدل‌ها با استفاده از یک زیرمجموعه کوچک‌تر از داده‌های برچسب‌گذاری شده تنظیم دقیق شدند.

۱-۴ ساختار پایان نامه

پایان‌نامه حاضر شامل پنج فصل اصلی است که هر یک به جنبه‌های مختلف پژوهش می‌پردازند. فصل اول به معرفی کلی پروژه، اهداف پژوهش اختصاص دارد. فصل دوم، به مفاهیم اصلی مورد استفاده در این تحقیق پرداخته می‌شود که شامل مباحثی مانند شبکه‌های عصبی پیچشی، یادگیری انتقالی و یادگیری خودناظارتی است. این فصل مفاهیم پایه را به تفصیل بیان می‌کند و چارچوب نظری تحقیق را شکل می‌دهد. فصل سوم به شرح پروژه و فرایندهای اجرایی آن می‌پردازد. این بخش، نحوه جمع‌آوری و پیش‌پردازش داده‌ها، انتخاب مدل‌ها و تنظیم آنها را مورد بررسی قرار می‌دهد. فصل چهارم به ارائه نتایج حاصل از مدل‌های یادگیری نظارتی و خودناظارتی اختصاص دارد که شامل ارزیابی‌های دقیق عملکرد هر مدل است. در نهایت، فصل پنجم به جمع‌بندی نتایج و تحلیل نهایی برای تحقیقات آینده می‌پردازد.

¹²Supervised Learning

مفاهیم

فصل دوم

۱-۲ شبکه‌های عصبی پیچشی

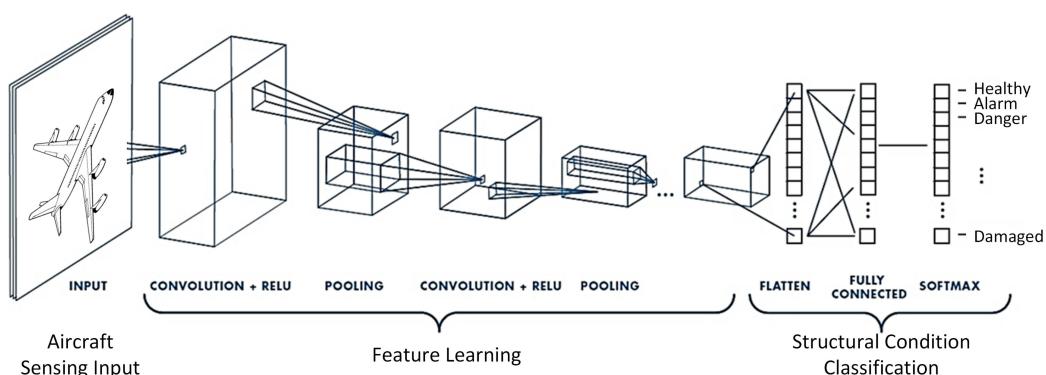
شبکه‌های عصبی پیچشی (CNNs)^۱ یک دسته از شبکه‌های عصبی عمیق هستند که به طور عمدۀ برای تحلیل داده‌های بصری مورد استفاده قرار می‌گیرند. این شبکه‌ها نقش مهمی در موفقیت وظایف بینایی کامپیوتر، مانند دسته‌بندی تصاویر، تشخیص اشیاء و بخش‌بندی تصاویر ایفا کرده‌اند.

۱-۱-۲ معرفی شبکه‌های عصبی پیچشی

شبکه‌های عصبی پیچشی به گونه‌ای طراحی شده‌اند که به صورت خودکار و تطبیقی سلسله مراتب مکانی ویژگی‌ها را از تصاویر ورودی یاد بگیرند. این شبکه‌ها از سیستم بصری انسان الهام گرفته شده‌اند، جایی که قشر بینایی شامل یک ترتیب پیچیده از سلول‌هایی است که مسئول شناسایی ویژگی‌های خاصی مانند لبه‌ها یا بافت‌ها در حرکت‌های بصری هستند. شبکه‌های عصبی پیچشی این فرآیند را با اعمال فیلترهای پیچشی روی داده‌های ورودی برای شناسایی ویژگی‌ها و الگوهای تقليید می‌کنند [۱۴].

۲-۱-۲ اجزای اصلی شبکه‌های عصبی پیچشی

یک شبکه عصبی پیچشی معمولاً از چندین نوع لایه تشکیل شده است که هر کدام وظیفه خاصی در عملیات شبکه دارند. این لایه‌ها شامل لایه‌های پیچشی^۲، توابع فعال‌سازی^۳، لایه‌های ادغام^۴ و لایه‌های کاملاً متصل می‌شوند.



شکل ۱-۲: شبکه عصبی پیچشی [۱]

• لایه‌های پیچشی

لایه پیچشی سنگ بنای شبکه‌های عصبی پیچشی است. این لایه شامل مجموعه‌ای از فیلترهای قابل یادگیری (یا هسته‌ها^۵) است که بر روی تصویر ورودی حرکت می‌کنند تا یک نقشه ویژگی تولید کنند. هر فیلتر ویژگی‌های خاصی مانند لبه‌ها یا گوش‌ها را شناسایی می‌کند. عملیات پیچش به صورت ریاضی به

¹Convolutional Neural Networks

²Convolutional Layer

³Activation Function

⁴Pooling

⁵Kernels

شکل زیر تعریف می‌شود:

$$\text{Feature Map} = f(W * X + b)$$

که در آن W فیلتر، X ورودی، $*$ نشان‌دهنده عملیات پیچش، b مقدار بایاس، و f تابع فعال‌سازی است که معمولاً^۶ یک واحد خطی تصحیح شده (ReLU) می‌باشد [۱۵]. این عملیات به شبکه امکان می‌دهد تا سلسله مراتب مکانی مختلفی از ویژگی‌ها را با افزایش عمق شبکه شناسایی کند.

• توابع فعال‌سازی

پس از عملیات پیچش، یک تابع فعال‌سازی به مدل اعمال می‌شود. تابع فعال‌سازی رایج در شبکه‌های عصبی پیچشی ReLU است که به صورت زیر تعریف می‌شود:

$$f(x) = \max(0, x)$$

ReLU به دلیل کاهش مشکل ناپدید شدن گرادیان‌ها^۷، ترجیح داده می‌شود و اجازه می‌دهد که مدل‌ها سریع‌تر و مؤثر‌تر یاد بگیرند [۱۶].

• لایه‌های ادغام (Pooling)

لایه‌های ادغام برای کاهش ابعاد مکانی نقشه‌های ویژگی^۸ استفاده می‌شوند. این عمل نمونه‌برداری به کاهش بار محاسباتی کمک می‌کند و همچنین از بیش برآش جلوگیری می‌کند. روش نمونه‌برداری بیشینه^۹، که بیشترین مقدار را از یک مجموعه مقادیر در یک پنجره ادغام می‌گیرد، رایج‌ترین روش است. این عملیات به شبکه‌های عصبی پیچشی امکان می‌دهد که به تغییرات کوچک در ورودی‌ها حساس نباشند [۱۷].

• لایه‌های کاملاً متصل

لایه‌های کاملاً متصل^۹ در مراحل نهایی یک شبکه عصبی پیچشی برای تصمیم‌گیری بر اساس ویژگی‌های استخراج‌شده توسط لایه‌های قبلی استفاده می‌شوند. این لایه‌ها هر نورون(Neuron) در یک لایه را به هر نورون در لایه بعدی متصل می‌کنند و اجازه می‌دهند تا سطح بالایی از انتزاع در ویژگی‌های یادگرفته شده ایجاد شود. لایه کاملاً متصل نهایی معمولاً از یک تابع فعال‌سازی softmax برای خروجی احتمالاتی در وظایف دسته‌بندی استفاده می‌کند [۱۸].

⁶Vanishing Gradient

⁷Spatial Size of the Representation

⁸Max pooling

⁹Fully Connected

۳-۱-۲ آموزش شبکه‌های عصبی پیچشی

آموزش شبکه‌های عصبی پیچشی شامل بهینه‌سازی یک تابع زیان^{۱۰} است که معمولاً برای وظایف دسته‌بندی از زیان آنتروپی متقاطع^{۱۱} استفاده می‌شود. این فرآیند با استفاده از پسانتشار^{۱۲} انجام می‌شود، جایی که گرادیان تابع زیان نسبت به هر وزن محاسبه شده و برای بهروزرسانی وزن‌ها از گرادیان نزولی استفاده می‌شود [۱۹]. تکنیک‌هایی مانند تنظیم نرخ یادگیری^{۱۳}، مومنتوم^{۱۴} و منظم‌سازی^{۱۵} (مانند کنار گذاشتن تصادفی نورون‌ها^{۱۶}) اغلب برای افزایش کارایی آموزش و جلوگیری از بیش‌برازش^{۱۷} به کار می‌روند.

۴-۱-۲ انواع شبکه‌های عصبی پیچشی

در طول سال‌ها، چندین معماری شبکه عصبی پیچشی برای بهبود عملکرد در وظایف خاص توسعه یافته‌اند.

LeNet-5 •

یکی از اولین شبکه‌های عصبی پیچشی که برای شناسایی اعداد طراحی شده است، شامل دو لایه پیچشی و سه لایه کاملاً متصل است [۱۸].

VGGNet •

به خاطر سادگی اش معروف است و از فیلترهای کوچک استفاده می‌کند و نشان می‌دهد که عمق شبکه برای عملکرد اهمیت دارد [۱۱].

ResNet •

اتصالات باقیمانده را معرفی کرد تا مشکل گرادیان‌های ناپدیدشونده در شبکه‌های عمیق را حل کند و امکان آموزش شبکه‌های بسیار عمیق با صدها لایه را فراهم کند [۲۰].

۲-۲ یادگیری انتقالی

یادگیری انتقالی^{۱۸} یک تکنیک در حوزه‌ی یادگیری ماشین و بهویژه یادگیری عمیق است که در آن مدلی که برای یک کار خاص آموزش دیده شده است، برای انجام وظایف دیگر نیز استفاده می‌شود. به عبارت دیگر، به جای این که از ابتدا مدل را برای هر کار جدید آموزش دهیم، از دانشی که مدل قبلاً در انجام وظایف مشابه کسب کرده است، استفاده می‌کنیم تا فرآیند یادگیری را تسريع و بهبود دهیم [۲۱].

¹⁰Loss Function

¹¹Cross-Entropy Loss

¹²Backpropagation

¹³Learning Rate Scheduling

¹⁴Momentum

¹⁵Regularization

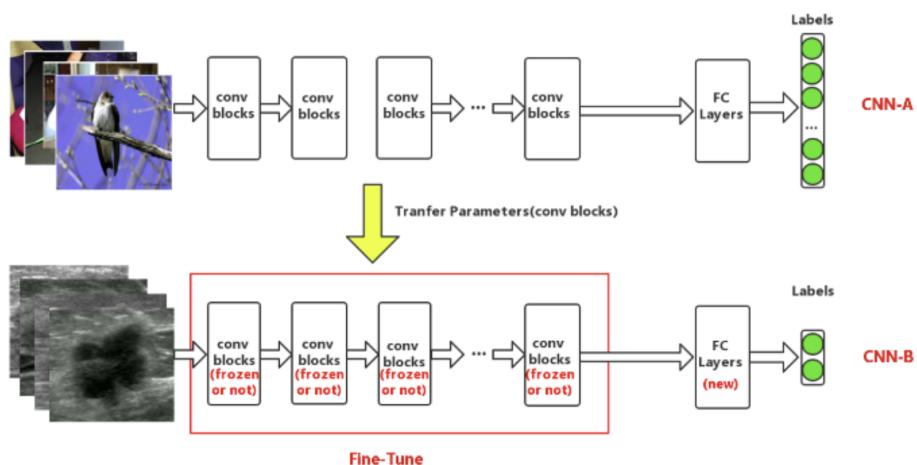
¹⁶Drop Out

¹⁷Overfitting

¹⁸Transfer Learning

۱-۲-۲ ایده کلی یادگیری انتقالی

ایده کلی یادگیری انتقالی این است که از دانشی که مدل برای حل یک وظیفه با حجم زیادی داده به دست آورده است برای وظیفه‌ی دیگری که این حجم داده را در دسترس ندارد، استفاده کنیم. در وظیفه‌های بینایی ماشین، شبکه‌های عصبی در لایه‌های اولیه به شناسایی لبه‌ها، در لایه‌های میانی به شناسایی اشکال و در لایه‌های نهایی به تشخیص ویژگی‌های خاص وظیفه‌ی مدنظر می‌پردازند. با استفاده از تکنیک یادگیری انتقالی، ما از لایه‌های اولیه و میانی وظیفه‌ی دیگری استفاده می‌کنیم و فقط لایه‌های نهایی را دوباره آموزش می‌دهیم تا به خروجی مدنظر در وظیفه‌ای برسیم که در حال کار روی آن هستیم.



شکل ۲-۲: یادگیری انتقالی [۲]

• مدل از پیش آموزش دیده: CNN-A

بخش بالای تصویر یک شبکه عصبی پیچشی (CNN) را نشان می‌دهد که بر روی یک مجموعه داده بزرگ (مانند ImageNet) که شامل میلیون‌ها تصویر برچسب‌خورده است) آموزش دیده است. این CNN که به آن CNN-A گفته می‌شود، شامل چندین لایه بلوك پیچشی است که قادر به استخراج ویژگی‌های مختلف از تصاویر هستند. این ویژگی‌ها از ویژگی‌های سطح پایین (مانند لبه‌ها، بافت‌ها) در لایه‌های ابتدایی تا ویژگی‌های پیچیده‌تر و انتزاعی‌تر (مانند اشکال و اجزای اشیاء) در لایه‌های عمیق‌تر متغیر هستند. در انتهای این شبکه، لایه‌های کاملاً متصل (FC^{۱۹}) وجود دارد که ویژگی‌های استخراج شده را به برچسب‌های خروجی خاص (دسته‌بندی‌ها) نگاشت می‌کنند. CNN-A به طور کامل آموزش دیده است، به این معنا که پارامترهای آن (وزن‌ها و بایاس‌ها در لایه‌های پیچشی و لایه‌های FC) برای وظیفه‌ای که برای آن آموزش دیده، بهینه شده‌اند.

• انتقال پارامترها:

پیکانی که از CNN-A به CNN-B اشاره می‌کند، انتقال پارامترها (وزن‌ها) از بلوك‌های پیچشی A به یک CNN جدید، با نام CNN-B را نشان می‌دهد. ایده این است که ویژگی‌های یادگرفته شده توسط

¹⁹Fully Connected

CNN-A به اندازه کافی عمومی هستند که می‌توانند برای یک وظیفه متفاوت، اما مرتبط، مجدد استفاده شوند. این اصل یادگیری انتقالی است.

• CNN-B (تطبیق دهنده دقیق برای یک وظیفه جدید):

CNN-B نشان‌دهنده مدل جدیدی است که شما برای یک وظیفه متفاوت اما مرتبط می‌سازید. برای مثال، اگر CNN-A بر روی تصاویر طبیعی آموزش دیده باشد، CNN-B ممکن است برای دسته‌بندی تصاویر پزشکی استفاده شود. بلوک‌های پیچشی از CNN-B به CNN-A منتقل می‌شوند. این بلوک‌ها می‌توانند ثابت^{۲۰} باشند (به این معنی که وزن‌های آن‌ها در طی آموزش CNN-B به روز نمی‌شوند) یا تنظیم دقیق^{۲۱} شوند (به این معنی که وزن‌های آن‌ها کمی به روزرسانی می‌شوند تا بهتر با وظیفه جدید سازگار شوند). لایه‌های کاملاً متصل (FC) در CNN-B معمولاً جدید هستند و از ابتدا آموزش می‌بینند، زیرا برچسب‌های خروجی برای وظیفه جدید با برچسب‌های CNN-A متفاوت است.

• تنظیم دقیق:

تنظیم دقیق فرآیند کمی تنظیم مجدد لایه‌های پیچشی از پیش آموزش دیده (در صورتی که ثابت نشده باشند) و آموزش لایه‌های کاملاً متصل جدید بر روی مجموعه داده جدید است. تنظیم دقیق به مدل این امکان را می‌دهد تا از دانش از پیش موجود در CNN-A بهره ببرد و در عین حال با ویژگی‌های خاص مجموعه داده جدید (مانند تصاویر پزشکی در این مورد) تطبیق یابد.

²⁰Frozen

²¹Fine Tune

۲-۲-۲ مزایای یادگیری انتقالی

مهم‌ترین مزیت‌های استفاده از تکنیک یادگیری انتقالی عبارتند از [۲۲]:

- کاهش زمان آموزش: زمان آموزش بهشت کاهش پیدا می‌کند.
- کاهش نیاز به داده: به حجم زیادی داده احتیاج نداریم.
- بهبود عملکرد: عملکرد شبکه‌ی عصبی در اغلب موارد بهتر خواهد بود.

۳-۲-۲ چالش‌های موجود در نبود داده‌های کافی

معمولًاً برای آموزش شبکه‌های عصبی به حجم زیادی داده نیاز داریم که این حجم از داده همیشه در دسترس نیست. اینجاست که یادگیری انتقالی می‌تواند کمک بزرگی باشد. آموزش شبکه از ابتدا به زمان زیادی نیاز دارد و گاهی ممکن است در وظیفه‌های پیچیده این موضوع هفته‌ها طول بکشد. استفاده از یادگیری انتقالی این زمان را کاهش می‌دهد؛ زیرا ما از شبکه‌ای که از قبل آموزش دیده است استفاده می‌کنیم.

۴-۲-۲ استفاده از مدل‌های پیش‌آموزش دیده

یکی از روش‌های استفاده از این تکنیک، بهره‌برداری از مدل‌های پیش‌آموزش دیده^{۲۲} است. یک مدل از پیش‌آموزش دیده، شبکه‌ای ذخیره‌شده است که قبلاً بر روی یک مجموعه داده بزرگ، معمولاً در یک وظیفه دسته‌بندی تصاویر در مقیاس بزرگ، آموزش دیده است. می‌توان از این مدل از پیش‌آموزش دیده به همان صورت استفاده کرد یا با استفاده از یادگیری انتقالی، این مدل را برای یک وظیفه خاص تنظیم دقیق کرد [۲۳].

۳-۲ یادگیری خودناظارتی

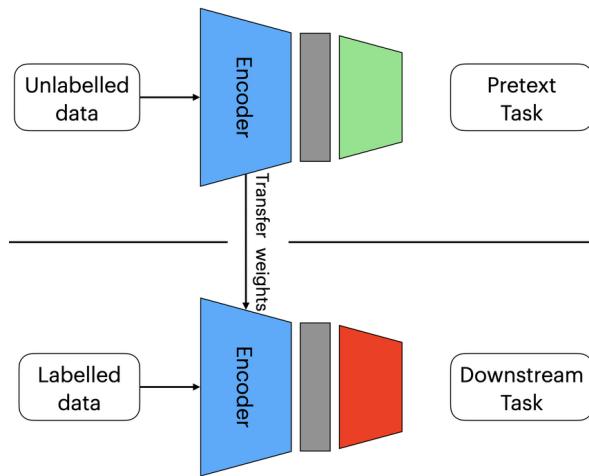
یادگیری خودناظارتی به عنوان یک تکنیک قدرتمند در حوزه یادگیری عمیق ظهره کرده است و به ویژه از سال 2020 به دلیل در دسترس بودن مجموعه داده‌های بزرگ و پیشرفت در توان محاسباتی مانند پردازنده‌های گرافیکی با حافظه بالا، رشد چشمگیری داشته است. با این حال، اصول و روش‌هایی که مبنای تکنیک‌های مدرن یادگیری خودناظارتی هستند، ریشه در روزهای اولیه یادگیری عمیق دارند.

۱-۳-۲ تعریف یادگیری خودناظارتی

یادگیری خودناظارتی یک روش یادگیری عمیق است که در آن یک مدل با استفاده از داده‌های بدون برچسب پیش‌آموزش داده می‌شود و برچسب‌های داده به طور خودکار تولید می‌شوند که در تکرارهای بعدی به عنوان حقایق مرجع استفاده می‌شوند. ایده اصلی یادگیری خودناظارتی این است که سیگنال‌های ناظارتی را از طریق درک داده‌های بدون برچسب در مرحله اول به صورت بدون ناظارت ایجاد کند. سپس، مدل از برچسب‌های داده با اطمینان بالا در میان برچسب‌های تولید شده برای آموزش مدل در تکرارهای بعدی استفاده می‌کند، مشابه با مدل

²²Pre-trained Models

یادگیری نظارت شده از طریق پس انتشار. تنها تفاوت این است که برچسب های داده ای که در هر تکرار به عنوان حقایق مرجع استفاده می شوند، تغییر می کنند [۲۴].



شکل ۳-۲: مراحل روش یادگیری خودنظرارتی [۳]

- **وظیفه پیش زمینه^{۲۳}**: وظیفه ای که برای پیش آموزش^{۲۴} استفاده می شود، به عنوان وظیفه پیش زمینه شناخته می شود. هدف وظیفه پیش زمینه (که به عنوان یک وظیفه نظارتی نیز شناخته می شود) هدایت مدل برای یادگیری بازنمایی های واسطه ای از داده ها است. این وظیفه در درک معنای ساختاری زیرین که برای وظایف عملیاتی بعدی مفید است، کمک می کند.

- **وظیفه پسین^{۲۵}**: پس از انجام پیش آموزش با استفاده از داده های بدون برچسب، مدل با استفاده از داده های برچسب گذاری شده کمتری تحت فرآیندی به نام تنظیم دقیق قرار می گیرد. در این مرحله، مدل با استفاده از بازنمایی های یادگرفته شده در مرحله پیش آموزش، برای یک وظیفه خاص بهینه سازی می شود. این فرآیند به مدل کمک می کند تا دقت و عملکرد خود را برای وظایف خاص بهبود دهد و دانش کسب شده از داده های بدون برچسب را به کار گیرد.

۲-۳-۲ ریشه یادگیری خودنظرارتی

ایده های پایه ای یادگیری خودنظرارتی از طریق روش هایی ایجاد شد که برای بهره برداری از ساختار درون داده ها طراحی شده بودند و به مدل ها اجازه می دادند بدون نیاز به داده های برچسب گذاری شده، بازنمایی هایی یاد بگیرند. رویکردهای اولیه یادگیری خودنظرارتی را می توان به چند استراتژی کلیدی تقسیم کرد [۲۵]:

- **بازیابی اطلاعات^{۲۶}**: این روش ها شامل حذف یا پوشاندن موقت بخش هایی از یک تصویر و آموزش مدل برای بازیابی اطلاعات از دست رفته است. یک مثال اولیه، یادگیری خودنظرارتی مبتنی بر رنگ آمیزی

²³Pretext tasks

²⁴Pretrain

²⁵Downstream Tasks

²⁶Information Restoration

است که در آن یک تصویر سیاه و سفید به عنوان ورودی استفاده می‌شود و مدل برای پیش‌بینی مقادیر اصلی RGB آموزش داده می‌شود، که به این ترتیب یادگیری ویژگی‌های اشیاء و مرزهای آن‌ها را تسهیل می‌کند^{[۲۶][۲۷]}. این ایده به تکنیک‌های پیچیده‌تری مانند رمزگذارهای خودکار با پوشاندن موقت بخش‌هایی از تصویر^[۲۸] تکامل یافته است، که در این روش، بخش‌هایی از تصویر به‌طور موقت پنهان می‌شوند و مدل برای پیش‌بینی و بازسازی این مناطق از دست رفته آموزش داده می‌شود، که معمولاً از ترانسفورمرها^[۲۹] برای انجام این کار استفاده می‌شود.

- **یادگیری زمینه مکانی^[۲۸]**: روش‌های یادگیری خودنظرارتی در این دسته مدل‌ها را برای درک روابط مکانی درون یک تصویر آموزش می‌دهند. برای مثال، روش‌هایی مانند RotNet^[۲۹] شامل چرخاندن یک تصویر و آموزش مدل برای پیش‌بینی میزان چرخش است، که در نتیجه، مدل در مورد جهت‌گیری اشیاء اطلاعات کسب می‌کند. رویکرد دیگری شامل تقسیم تصویر به قطعات و آموزش مدل برای پیش‌بینی موقعیت نسبی این قطعات است، که درک مدل از زمینه مکانی را افزایش می‌دهد.

- **گروه‌بندی تصاویر مشابه^[۲۹]**: این رویکرد شامل خوشبندی تصاویر مشابه با هم است که به مدل‌ها اجازه می‌دهد ویژگی‌های غنی را از ساختار ذاتی داده‌ها یاد بگیرند. تکنیک‌هایی مانند خوشبندی عمیق^[۳۰]، بین اختصاص دادن برچسب‌ها به تصاویر با استفاده از الگوریتم k-means در فضای ویژگی و به‌روزرسانی مدل بر اساس این برچسب‌ها، جایجاً می‌شوند^[۳۰]. این روش با تکنیک‌هایی مانند به‌روزرسانی میانگین شیفت و برچسب‌گذاری مجازی، که عملکرد خوشبندی را بهبود می‌بخشند، بیشتر توسعه یافته است.

²⁷Transformers

²⁸Learning Spatial Context

²⁹Grouping Similar Images

³⁰Deep Clustering

- عدم تغییر در چند نمایی: ^{۳۱} روش‌های مدرن یادگیری خودنظراتی، به ویژه تکنیک‌های یادگیری متضاد، بر اطمینان از عدم تغییر بازنمایی‌های یک ورودی تحت تغییرات مختلف متمرکز هستند. به عنوان مثال، SimCLR، تغییرات مختلفی را به یک تصویر اعمال می‌کند و مدل را آموزش می‌دهد تا بازنمایی‌های مشابهی برای این نماهای مختلف تولید کند، که به این ترتیب عدم تغییر در تقویت داده‌ها را تضمین می‌کند.

۳-۳-۲ تکامل خانواده‌های یادگیری خودنظراتی

همانطور که یادگیری خودنظراتی تکامل یافته است، روش‌های معاصر به چند خانواده تقسیم شده‌اند که هر یک بر اساس ایده‌های پایه‌ای فوق توسعه یافته‌اند [۲۵]:

- خانواده یادگیری متریک عمیق (SimCLR^{۳۶}) (DML^{۳۲}): این خانواده بر یادگیری بازنمایی‌های بصری از طریق تشویق به شیاهت بین نسخه‌های تغییر یافته یک ورودی متمرکز است. ابزار اصلی این روش، زیان متضاد^{۳۳} است که با معرفی مکانیزم‌هایی مانند زیان^{۳۴} [۲۱] InfoNCE، که هدف آن افزایش اطلاعات متقابل بین بازنمایی‌های نماهای مختلف است، بیشتر پالایش یافته است.

• خانواده خود-تقطیر (BYOL^{۳۵}) (DINO^{۳۴}, SimSiam^{۳۲}, BYOL^{۳۶}):

این روش‌ها شامل آموزش دو شبکه هستند، که یکی از آن‌ها (دانش‌آموز) می‌آموزد که خروجی‌های شبکه دیگر (معلم) را پیش‌بینی کند. تکنیک‌هایی مانند BYOL از طریق استفاده از مکانیزم‌هایی مانند رمزگذاران مونتوم و شبکه‌های پیش‌بینی کننده، از فروپاشی بازنمایی‌های آموخته شده جلوگیری می‌کند و اطمینان می‌دهند که شبکه دانش‌آموز ویژگی‌های معناداری را یاد می‌گیرد.

- خانواده تحلیل همبستگی کانونی (Barlow Twins^{۳۷}, VICReg^{۳۵}, CCA^{۳۸}): روش‌های این خانواده بر افزایش همبستگی بین بازنمایی‌های نماهای مختلف یک تصویر متمرکز هستند. هدف این است که بازنمایی‌هایی آموخته شوند که هم در برابر تغییرات پایدار باشند و هم در ابعاد مختلف متنوع باشند، که از تکرار جلوگیری کرده و اطمینان می‌دهد که ویژگی‌های غنی کدگذاری می‌شوند.

- مدل‌سازی تصویر پوشانده شده (MAE^{۲۸}, BEiT^{۳۶}): این روش‌ها که از مدل‌سازی زبان پوشانده شده در پردازش زبان طبیعی (NLP)^{۳۹} الهام گرفته‌اند، شامل پوشاندن بخش‌هایی از یک تصویر و آموزش مدل برای بازسازی قسمت‌های از دست رفته هستند. این رویکرد در یادگیری بازنمایی‌های مقاوم

³¹ Multi-view invariance

³² Deep Metric Learning Family

³³ Contrastive Loss

³⁴ Information Noise-Contrastive Estimation

³⁵ Self-Distillation Family

³⁶ Bootstrap Your Own Latent

³⁷ Canonical Correlation Analysis

³⁸ Masked Image Modeling

³⁹ Natural language processing

برای وظایف پسین مؤثر بوده و از معماری ترانسفورمر برای مدلسازی وابستگی‌های پیچیده درون داده‌ها استفاده می‌کند.

فصل سوم

شرح پروژه

۱-۳ جمع آوری داده

۱-۱-۳ معرفی مجموعه داده‌ها

در این مطالعه، از دو مجموعه داده مجزا اما مکمل استفاده شد، مجموعه داده تصاویر رادیوگرافی قفسه سینه سل [۱۲] (Tuberculosis (TB) Chest X-ray Database) و مجموعه داده تصاویر رادیوگرافی قفسه سینه ذات‌الریه [۱۳] (Chest X-Ray Images (Pneumonia)). این مجموعه داده‌ها بر اساس ارتباطشان با هدف تحقیق، که مقایسه عملکرد روش‌های مختلف یادگیری نظارت شده و یادگیری خودناظارتی در تشخیص سل و ذات‌الریه از تصاویر رادیوگرافی قفسه سینه است، انتخاب شدند. ترکیب این مجموعه داده‌ها امکان بررسی گسترده‌تری از قابلیت‌های مدل‌های آموزش‌دیده را فراهم کرد و همچنین یک مجموعه داده جامع‌تر برای کار فراهم آورد.

- مجموعه داده تصاویر رادیوگرافی قفسه سینه سل:** مجموعه داده تصاویر رادیوگرافی قفسه سینه سل توسط یک همکاری مشترک بین محققان دانشگاه قطر، داکا، بنگلادش و همکاران آن‌ها از مالزی، با همکاری متخصصان پزشکی از شرکت پزشکی بنگلادش ایجاد شد. این مجموعه داده تصاویر به طور خاص برای کمک به توسعه ابزارهای تشخیصی خودکار برای سل طراحی شده است و مجموعه بزرگی از تصاویر رادیوگرافی قفسه سینه را با تمرکز خاص بر موارد مثبت سل ارائه می‌دهد. این مجموعه داده شامل مجموعاً ۳۵۰۰ تصویر عادی است. تصاویر مثبت سل، ۷۰۰ تصویر به صورت عمومی در دسترس است. هر تصویر رادیوگرافی در این مجموعه داده به دقت انتخاب شده تا از کیفیت بالا اطمینان حاصل شود و این مجموعه داده توسط متخصصان پزشکی برچسب‌گذاری شده است که اطمینان از صحت برچسب‌ها را تضمین می‌کند. این امر در وظایف تصویربرداری پزشکی که کیفیت و دقت داده‌ها مستقیماً بر عملکرد مدل‌های آموزش‌دیده تأثیر می‌گذارد، بسیار حیاتی است.

- مجموعه داده تصاویر رادیوگرافی قفسه سینه ذات‌الریه مجموعه داده دوم استفاده شده در این مطالعه، مجموعه داده تصاویر رادیوگرافی قفسه سینه ذات‌الریه است.** این مجموعه داده به سه پوشه اصلی تقسیم شده است: آموزش، آزمایش و اعتبارسنجی، که هر پوشه شامل زیرپوشه‌هایی برای هر دسته از تصاویر (ذات‌الریه/عادی) است. این مجموعه داده شامل مجموعاً ۵۸۶۳ تصویر رادیوگرافی قفسه سینه است. این تصاویر از یک گروه بیماران کودکان با سنین بین یک تا پنج سال از مرکز پزشکی زنان و کودکان گوانگزو به دست آمده است. تمامی تصویربرداری‌های قفسه سینه به عنوان بخشی از مراقبت‌های بالینی روزمره انجام شده است. این مجموعه داده به ویژه ارزشمند است زیرا تعداد قابل توجهی از موارد کودکان را ارائه می‌دهد که به تنوع مجموعه داده کلی مورد استفاده در این مطالعه اضافه می‌کند. قبل از اینکه در مجموعه داده گنجانده شوند، تمامی تصاویر تحت فرآیند کنترل کیفیت دقیق قرار گرفتند. اسکن‌های با کیفیت پایین یا غیرقابل خواندن حذف شدند تا اطمینان حاصل شود که مجموعه داده فقط شامل تصاویر با کیفیت بالا است که برای آموزش مدل‌های یادگیری ماشین مناسب هستند. علاوه بر این، تشخیص‌ها برای هر تصویر توسط دو پزشک متخصص ارزیابی شده است. برای کاهش احتمال خطاهای ارزیابی، مجموعه ارزیابی نیز توسط یک متخصص سوم بررسی شده است. این سطح از کنترل کیفیت به اطمینان از قابلیت اطمینان مجموعه داده و دقت مدل‌های حاصل کمک می‌کند.

۲-۳ پیش‌پردازش داده‌ها و ایجاد مجموعه داده

پس از جمع‌آوری دو مجموعه داده، مرحله بعدی شامل پیش‌پردازش و ترکیب آن‌ها برای ایجاد یک مجموعه داده یکپارچه جهت آموزش و ارزیابی بود. مراحل پیش‌پردازش شامل تغییر نام پوشش‌ها، تقسیم مجموعه داده‌ها به مجموعه‌های آموزش، اعتبارسنجی و آزمایش و رفع هرگونه مشکلات عدم تعادل دسته‌ها بود.

۱-۲-۳ تغییر نام و ساختاردهی مجموعه داده

در ابتدا، مجموعه داده رادیوگرافی قفسه سینه سل نیاز به برخی سازماندهی مجدد داشت تا با ساختار مجموعه داده تصاویر رادیوگرافی قفسه سینه ذات‌الریه همخوانی داشته باشد. این امر شامل تغییر نام پوشش‌ها برای استانداردسازی برچسب‌گذاری دسته‌ها بود. به طور خاص، پوشش‌های حاوی تصاویر عادی و سل به ترتیب به 'NORMAL' و 'TUBERCULOSIS' تغییر نام دادند. علاوه بر این، هرگونه فایل غیرضروری مانند صفحات گسترده اطلاعات متابه و فایل‌های راهنمای از مجموعه داده حذف شدند تا ساختار مجموعه داده ساده‌تر شود و تمرکز فقط بر روی تصاویر باشد.

۲-۲-۳ تقسیم مجموعه داده

سپس هر دو مجموعه داده به سه دسته^۱ تقسیم شدند: آموزش^۲، اعتبارسنجی^۳ و آزمایش^۴. تقسیم به نسبت ۱۵-۱۵-۷۰ انجام شد، به این معنا که ۷۰ درصد از داده‌ها برای آموزش، ۱۵ درصد برای اعتبارسنجی و ۱۵ درصد برای آزمایش تخصیص داده شد. این یک روش رایج در یادگیری ماشین است تا اطمینان حاصل شود که مدل به اندازه کافی داده برای یادگیری دارد، در حالی که داده‌های جداگانه‌ای نیز برای اعتبارسنجی و آزمایش عملکرد مدل وجود دارد.

برای اطمینان از اینکه تقسیم به درستی انجام شده است، هر دسته در مجموعه داده رادیوگرافی قفسه سینه سل ('NORMAL' و 'TUBERCULOSIS') به صورت جداگانه تقسیم شد. تصاویر در هر دسته به صورت تصادفی مخلوط شدند قبل از تقسیم، تا اطمینان حاصل شود که داده‌ها در هر زیرمجموعه نماینده کل دسته هستند. این مخلوطسازی تصادفی به جلوگیری از هرگونه سوگیری کمک می‌کند که ممکن است در صورت تقسیم تصاویر به صورت متوالی یا به روی غیرتصادفی ایجاد شود.

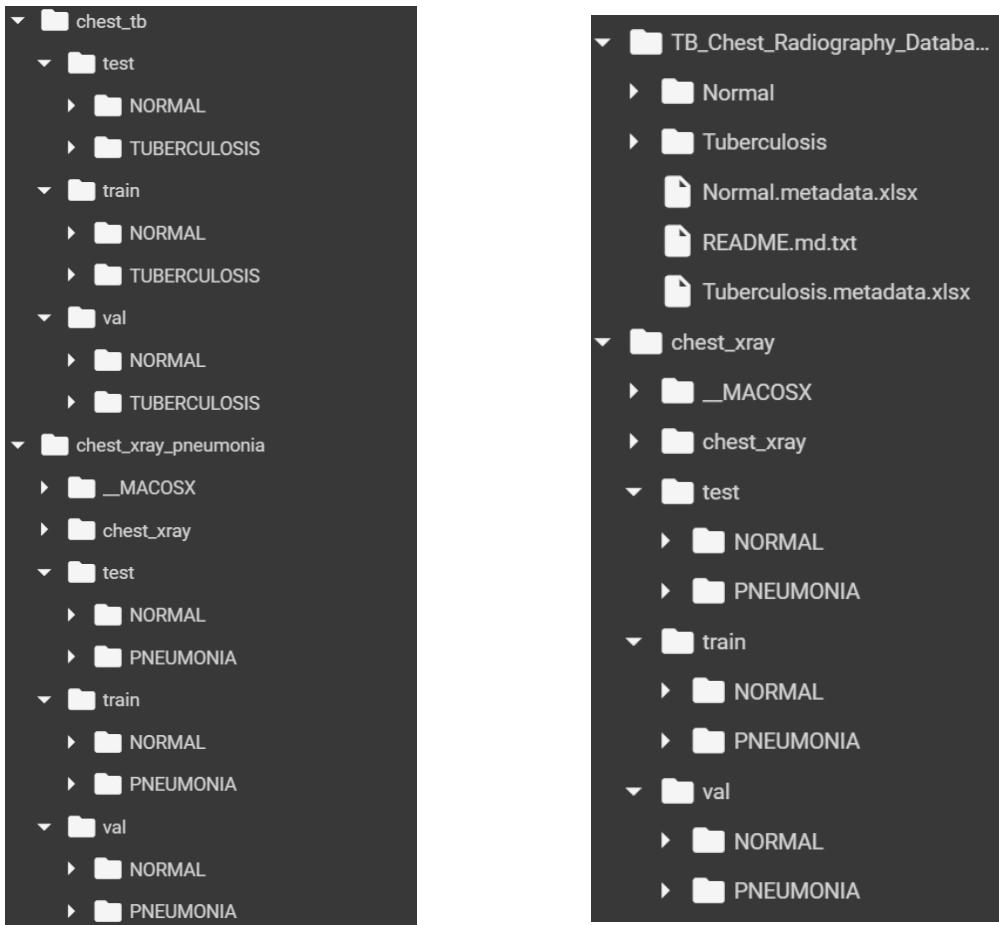
همین فرآیند تقسیم برای مجموعه داده تصاویر رادیوگرافی قفسه سینه ذات‌الریه نیز اعمال شد، با این تفاوت که تصاویر به دسته‌های 'Normal' و 'Pneumonia' قبل از تقسیم، دسته‌بندی شدند.

¹Class

²Training

³Validation

⁴Testing



شکل ۳-۲: مجموعه داده قبل از پیش پردازش

شکل ۳-۳: مجموعه داده قبل از پیش پردازش

۳-۲-۳ ترکیب مجموعه داده‌ها

پس از اینکه هر دو مجموعه داده به مجموعه‌های آموزش، اعتبارسنجی و آزمایش تقسیم شدند، آن‌ها برای ایجاد یک مجموعه داده واحد با سه دسته: 'NORMAL', 'TUBERCULOSIS' و 'PNEUMONIA' ترکیب شدند. این ترکیب برای مطالعه ضروری بود زیرا به مدل امکان داد تا بر روی یک مجموعه داده متنوع‌تر که شامل تصاویر رادیوگرافی قفسه سینه عادی، تصاویر مثبت سل و موارد ذات‌الریه بود، آموزش بینند.

ترکیب مجموعه داده‌ها با استفاده از تابع `ConcatDataset` انجام شد که امکان ادغام بی‌نقص مجموعه داده‌های متعدد را فراهم می‌کند. با ادغام مجموعه داده‌ها، مدل می‌توانست از ویژگی‌ها و خصوصیات مختلف موجود در تصاویر هر دو مجموعه داده بهره‌برداری کند، که به طور بالقوه منجر به تعمیم بهتر و بهبود عملکرد در داده‌های دیده‌نشده می‌شد.

۴-۲-۳ رفع عدم تعادل دسته‌ها

پس از ترکیب مجموعه داده‌ها، یکی از مسائل مهمی که باید به آن پرداخته می‌شد، عدم تعادل دسته‌ها بود. در یادگیری ماشین، عدم تعادل دسته‌ها زمانی رخ می‌دهد که یک دسته با تعداد قابل توجهی بیشتری نمونه نسبت به دسته‌های دیگر نمایانده شود. این امر می‌تواند منجر به ایجاد مدلی شود که بر روی دسته اکثریت عملکرد خوبی دارد اما در دسته‌های اقلیت عملکرد ضعیفی از خود نشان می‌دهد.

برای حل این مسئله، تعداد نمونه‌ها برای هر دسته ابتدا مورد تحلیل قرار گرفت. تحلیل نشان داد که مجموعه داده ترکیبی نامتعادل بود، با تعداد بیشتری از تصاویر عادی نسبت به دسته‌های سل و ذاتالریه. برای رفع این عدم تعادل، چندین مرحله انجام شد:

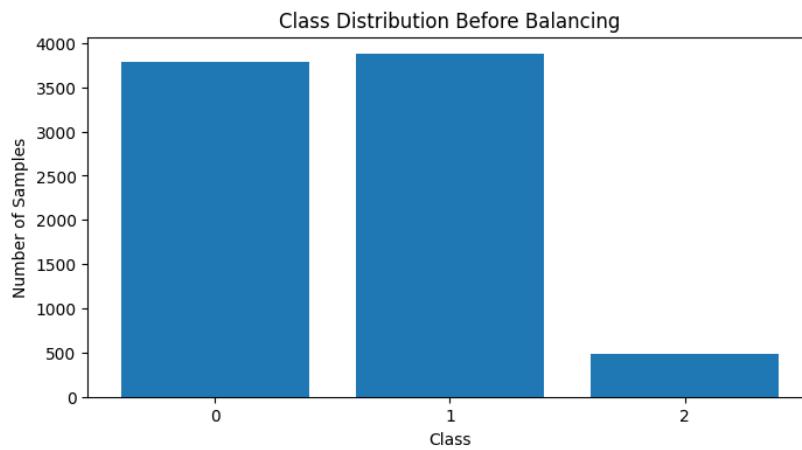
- **وزن دهنده دسته‌ها:** وزن‌های دسته به عنوان معکوس فراوانی دسته‌ها محاسبه شد. این بدان معناست که دسته‌هایی با تعداد نمونه کمتر وزن بیشتری دریافت کردند و دسته‌هایی با تعداد بیشتر وزن کمتری داشتند. این وزن‌ها سپس برای تنظیم تابع زیان در طول آموزش مدل استفاده شدند تا اطمینان حاصل شود که مدل به دسته‌های اقلیت توجه بیشتری می‌کند.

- **نمونه‌برداری تصادفی وزنی:** یک `WeightedRandomSampler` در طول آموزش استفاده شد تا اطمینان حاصل شود که هر زیرمجموعه کوچک داده^۵ ارائه شده به مدل دارای نمایندگی متعادلی از هر دسته است. این نمونه‌بردار با اختصاص احتمال بالاتری برای انتخاب به نمونه‌های دسته‌های اقلیت، به طور مؤثر مجموعه داده را در طول آموزش متعادل می‌سازد.

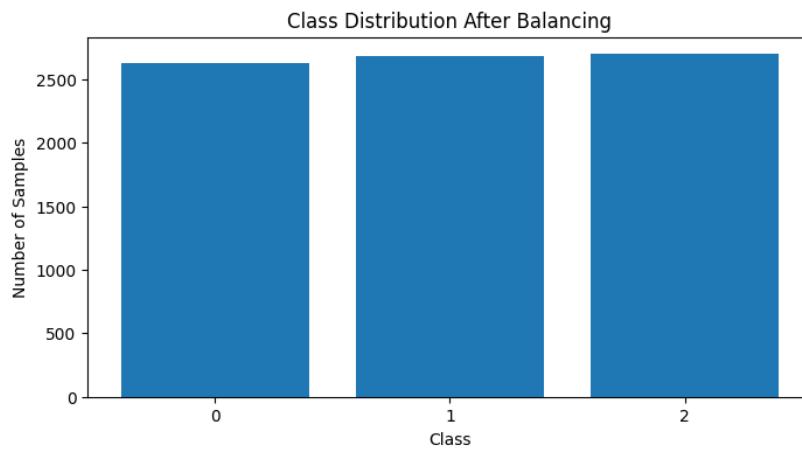
- **تحلیل نمونه‌برداری شبیه‌سازی شده:** برای تأیید اثربخشی تکنیک‌های متعادل‌سازی، یک تحلیل نمونه‌برداری شبیه‌سازی شده انجام شد. این شامل نمونه‌برداری از تعداد زیادی داده با استفاده از `WeightedRandomSampler` و مقایسه توزیع دسته‌ها قبل و بعد از اعمال تکنیک‌های متعادل‌سازی بود. نتایج نشان داد که تعادل دسته به طور قابل توجهی بهبود یافته است، که نشان می‌دهد مدل به مجموعه داده متعادل‌تری در طول آموزش دسترسی خواهد داشت.

این نمودارها درکی واضح و شهودی از چگونگی تغییرات اعمال شده بر روی مجموعه داده ارائه دادند، و اطمینان حاصل شد که فرآیند آموزش مدل منصفانه‌تر و مؤثرتر خواهد بود.

⁵Mini-Batch



شکل ۳-۳: دسته‌ها قبل از متعادل‌سازی



شکل ۴-۳: دسته‌ها بعد از متعادل‌سازی

۵-۲-۳ نمای کلی از مجموعه داده نهایی

مجموعه داده نهایی مورد استفاده در این مطالعه شامل سه دسته 'PNEUMONIA', 'NORMAL' و 'TUBERCULOSIS' بود. این مجموعه داده به مجموعه‌های آموزش، اعتبارسنجی و آزمایش تقسیم شد که هر مجموعه دارای نمایندگی متعادلی از سه دسته بود. مجموعه داده با دقت انتخاب شده بود تا از کیفیت بالا و قابلیت اطمینان حاصل شود و مراحل پیش‌پردازش دقیق برای رفع هرگونه مشکلات مربوط به عدم تعادل دسته‌ها انجام شده بود. این مجموعه داده مبنای مراحل بعدی تحقیق را تشکیل داد، که در آن روش‌های مختلف یادگیری نظارت شده و یادگیری خودناظارتی به کار گرفته شدند تا مدل‌های قادر به تشخیص دقیق سل و ذات‌الریه از تصاویر رادیوگرافی قفسه سینه آموزش داده شوند. با ترکیب و متعادل‌سازی این مجموعه داده‌ها، تحقیق به دنبال توسعه مدل‌هایی بود که نه تنها عملکرد خوبی داشتند بلکه به طور مؤثر به داده‌های جدید و دیده‌نشده تعمیم پیدا می‌کردند. به طور کلی، انتخاب دقیق، پیش‌پردازش و متعادل‌سازی مجموعه داده‌ها گام‌های حیاتی در تضمین موفقیت

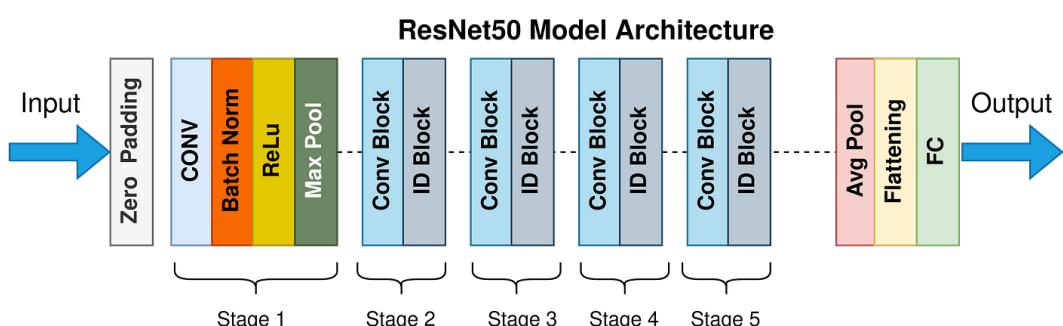
مدل‌های یادگیری ماشین توسعه یافته در این مطالعه بودند. توجه به جزئیات در این فرآیندها نشان‌دهنده اهمیت داده‌های با کیفیت بالا در دستیابی به نتایج دقیق و قابل اطمینان در تحقیق تصویربرداری پزشکی است.

۳-۳ یادگیری ناظارت شده

در این پژوهه، دو مدل یادگیری عمیق مختلف، ResNet50 و VGG16 برای دسته‌بندی تصاویر پزشکی به سه دسته مورد استفاده قرار گرفتند. این فرآیند شامل انتخاب و استفاده از مدل‌های از پیش‌آموزش دیده از طریق یادگیری انتقالی، تنظیم دقیق، و ارزیابی دقیق بود. در ادامه، توضیحی جامع از مراحل مختلف، با تمرکز بر استفاده از هر دو مدل ResNet50 و VGG16 ارائه می‌شود.

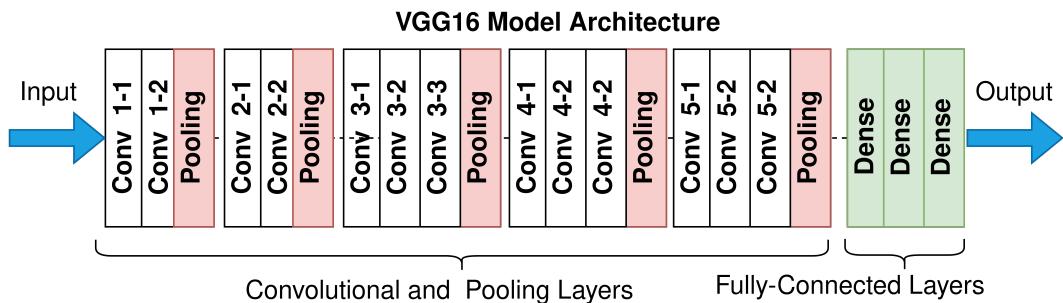
۱-۳-۳ مرحله ۱: انتخاب مدل و یادگیری انتقالی

ResNet50: یک شبکه عمیق با ۵۰ لایه است که یادگیری باقیمانده (Residual Learning) را به کار می‌گیرد، که به آن امکان می‌دهد تا شبکه‌های بسیار عمیق را با کاهش مشکل ناپدید شدن گرادیان‌ها به طور مؤثر آموزش دهد. این مدل به دلیل توانایی اش در یادگیری الگوهای پیچیده در تصاویر به دلیل عمق زیاد آن انتخاب شد.



[۴]ResNet50: معماری شبکه

VGG16: VGG16 یک معماری محبوب دیگر در CNN است که به دلیل سادگی و اثربخشی اش در وظایف دسته‌بندی تصویر شناخته می‌شود. این مدل شامل ۱۶ لایه است و به دلیل استفاده از فیلترهای کوچک 3×3 و تعداد زیاد لایه‌ها، می‌تواند جزئیات دقیق‌تری را در تصاویر به خوبی شناسایی کند. VGG16 به دلیل عملکرد اثبات‌شده‌اش در موقعیت‌هایی که اطلاعات مکانی دقیق حیاتی است، انتخاب شد.



شکل ۳-۶: معماری شبکه [۵]VGG16

یادگیری انتقالی: به جای آموزش مدل‌ها از ابتدا، از یادگیری انتقالی استفاده کردیم که از وزن‌های از پیش‌آموزش‌دیده مدل‌ها که بر روی مجموعه داده ImageNet آموزش دیده است، بهره می‌برد. مجموعه داده ImageNet شامل میلیون‌ها تصویر در 10^6 دسته است. این مدل‌های از پیش‌آموزش‌دیده یک نقطه شروع قوی برای وظایف دیگر دسته‌بندی تصاویر است.

بارگذاری مدل از پیش‌آموزش‌دیده: مدل‌ها با وزن‌های پیش‌فرض آموزش‌دیده بر روی مجموعه داده ImageNet بارگذاری شدن. این امر به مدل‌ها اجازه داد تا از ویژگی‌های یادگرفته‌شده از ImageNet استفاده کنند که به اندازه کافی عمومی هستند تا در دسته‌بندی تصاویر پزشکی مفید باشند.

تغییر لایه نهایی: از آنجا که وظیفه ما شامل دسته‌بندی تصاویر به سه دسته است، لایه کاملاً متصل نهایی مدل‌ها که در ابتدا 10^6 دسته را خروجی می‌داد، تغییر داده شده است. این لایه با یک لایه کاملاً متصل جدید که سه دسته را خروجی می‌دهد، جایگزین شده است. این تغییر اطمینان حاصل کرد که معماری مدل با نیازهای وظیفه دسته‌بندی خاص ما همخوانی دارد.

۲-۳ مرحله ۲: تنظیم دقیق مدل

در تنظیم دقیق به مدل از پیش‌آموزش‌دیده اجازه داده می‌شود تا وزن‌های خود را بر اساس مجموعه داده ما تنظیم کند. این مرحله به ویژه در یادگیری انتقالی مؤثر است، زیرا مدل را قادر می‌سازد تا به مجموعه داده جدید تطبیق یابد در حالی که ویژگی‌های قدرتمند یادگرفته‌شده از مجموعه داده بزرگ ImageNet را حفظ می‌کند. در این مرحله، از تابع زیان CrossEntropyLoss و تابع بهینه‌ساز Adam با نرخ یادگیری $1e-3$ استفاده شده است.

۳-۳-۳ مرحله ۳: ارزیابی مدل و توقف زودهنگام

در طول فرآیند آموزش، عملکرد مدل بر روی مجموعه اعتبارسنجی ارزیابی شده و برای جلوگیری از بیشبرازش، از توقف زودهنگام استفاده شد. توقف زودهنگام^۶ یک روش است که در آن فرآیند آموزش زمانی متوقف می‌شود که عملکرد مدل بر روی مجموعه اعتبارسنجی برای تعداد مشخصی از دورها بهبود نمی‌یابد. در این مورد، آموزش در صورتی متوقف شد که به مدت پنج دوره متوالی هیچ بهبودی در زیان اعتبارسنجی^۷ مشاهده نشد.

۴-۳-۳ مرحله ۴: جمع‌آوری معیارها و نظارت بر عملکرد

در طول مراحل آموزش و اعتبارسنجی، چندین معیار ریدیابی شد که شامل موارد زیر بود:

- زیان^۸: تابع زیان آنتروپی که خطا بین احتمالات دسته پیش‌بینی شده و برچسب‌های واقعی دسته را اندازه‌گیری می‌کند.
- دقت^۹: درصد تصاویر به درستی دسته‌بندی شده.
- فراخوانی^{۱۰} و امتیاز F1^{۱۱}: این معیارها برای ارائه درک دقیق‌تری از عملکرد مدل، به ویژه در برخورد با عدم تعادل دسته‌ها محاسبه شدند.

۵-۳-۳ مرحله ۵: تست مدل

پس از اتمام فرآیند آموزش، مدل بر روی یک مجموعه داده تست جداگانه ارزیابی شد. این مجموعه داده در طول آموزش یا اعتبارسنجی استفاده نشده بود تا یک ارزیابی بی‌طرفانه از عملکرد مدل فراهم شود.

۴-۳ یادگیری خودناظارتی

در این پژوهش، جهت مقایسه روش‌های یادگیری خودناظارتی با یادگیری نظارت‌شده، از دو روش SimCLR و BarlowTwins با شبکه ResNet50 به عنوان رمزگذار^{۱۲} استفاده شده است. جهت ارزیابی این دو روش، از دو

⁶Early Stopping

⁷Validation Loss

⁸Loss

⁹Accuracy

¹⁰Recall

¹¹F1-Score

¹²Encoder

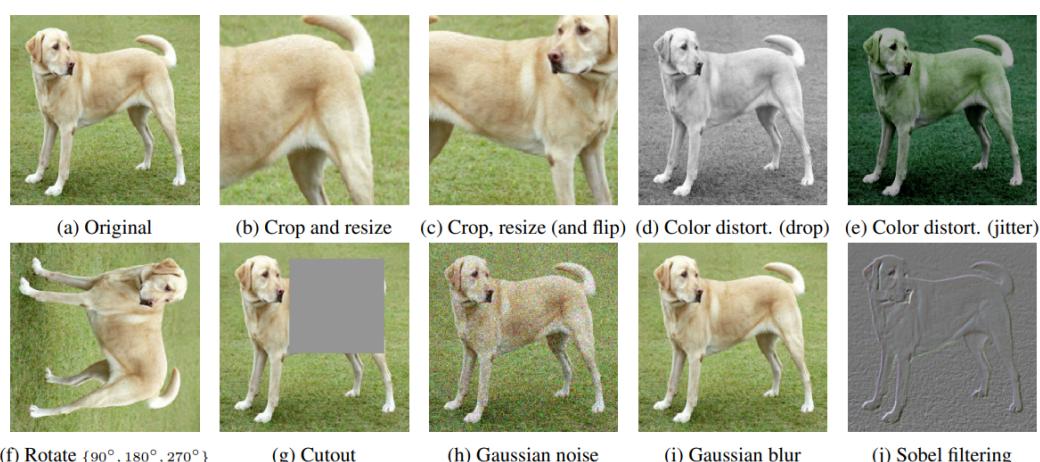
راهکار ارزیابی خطی^{۱۳} و تنظیم دقیق استفاده شده است. این رویکرد به ما امکان می‌دهد تا بررسی کنیم که تنظیم دقیق کل شبکه در مقابل آموزش تنها لایه نهایی چگونه بر عملکرد تأثیر می‌گذارد.

SimCLR ۱-۴-۳

SimCLR (چارچوب ساده برای یادگیری تمایزی بازنمایی‌های بصری^{۱۴}) یک روش یادگیری خودناظارتی است که برای یادگیری بازنمایی‌های بصری از داده‌های بدون برچسب طراحی شده است. ایده اصلی در SimCLR این است که از افزایش داده و یادگیری تمایزی استفاده کند تا مدلی را آموزش دهد که بتواند بین نمونه‌های مختلف (یا تصاویر) تمایز قائل شود و نمونه‌های مشابه (افزایش‌های مختلف از یک تصویر) را در فضای بازنمایی به هم نزدیک‌تر کند.

مفاهیم کلیدی SimCLR

- افزایش داده^{۱۵}: نقش مهمی در SimCLR ایفا می‌کند. با ایجاد نماهای مختلف از یک تصویر، مدل یاد می‌گیرد که نسبت به انواع تغییرات مقاوم باشد. افزایش‌های استانداردی که در SimCLR استفاده می‌شود.



شکل ۷-۳: افزایش‌های تصویر در SimCLR [۶]

این افزایش‌ها به مدل کمک می‌کنند تا بازنمایی‌های مقاومی را یاد بگیرد که نسبت به این تغییرات نامتفاوت باشند، این عمل برای یادگیری ویژگی‌های قابل تعمیم بسیار مهم است. در پیاده‌سازی این پروژه از برش GrayScale و Change Contrast (Crop)، تغییرات رنگ (Color Jitter)، تغییرات افقی (Flip)، تغییرات افقی (Resize)، و تغییر اندازه (Scale).

¹³Linear Evaluation

¹⁴A Simple Framework for Contrastive Learning of Visual Representations

¹⁵Data Augmentation

تصادفی و Gaussian Blur استفاده کردیم و به طور کلی در SimCLR دو افزایش داده به صورت تصادفی انتخاب شده و بر روی تصویر اعمال می‌شوند که این عمل در این پژوهه نیز پیاده‌سازی شده است.

- **رمزنگار شبکه عصبی:** تصاویر افزوده شده از طریق یک رمزنگار شبکه عصبی عبور می‌کند که در این پژوهه یک شبکه عصبی پیچشی عمیق (مانند ResNet50) است. این رمزنگار ویژگی‌ها را از تصاویر استخراج می‌کند. خروجی از رمزنگار یک بردار ویژگی با ابعاد بالا است که اغلب به عنوان تعبیه^{۱۶} نامیده می‌شود.

- **سر پیش‌بینی^{۱۷}:** تعبیه‌های تولید شده توسط رمزنگار از طریق یک سر پیش‌بینی که معمولاً یک پرسپترون چند لایه (MLP^{۱۸}) است، عبور می‌کند. سر پیش‌بینی، بردار ویژگی^{۱۹} با ابعاد بالا به فضای با بعد کمتر می‌برد که در آن از تابع زیان تمایزی^{۲۰} اعمال می‌شود. استفاده از سر پیش‌بینی به یادگیری بازنمایی‌های بهتر کمک می‌کند، زیرا فضای تعبیه برای یادگیری تمایزی از فضای مورد استفاده برای وظایف پسین جدا می‌شود.

- **تابع زیان تمایزی (NT-Xent Loss)^{۲۱}:** هسته SimCLR تابع زیان تمایزی است که به عنوان از دست دادن آنتروپی مقاطع مقیاس‌شده با دما (NT-Xent) شناخته می‌شود. هدف این تابع زیان این است که نمونه‌های مشابه (جفت‌های مثبت) را در فضای تعبیه به هم نزدیک کند در حالی که نمونه‌های غیرمشابه (جفت‌های منفی) را از هم دور کند.

تابع زیان به صورت زیر به صورت ریاضی تعریف می‌شود:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^N \mathbb{1}_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$

که در آن:

- $\text{sim}(\mathbf{z}_i, \mathbf{z}_j)$ شباهت کسینوسی بین دو بردار ویژگی \mathbf{z}_i و \mathbf{z}_j است.
- τ یک پارامتر دما است که تندی توزیع را کنترل می‌کند.
- N اندازه دسته^{۲۲} است، بنابراین در کل $2N$ نمونه افزوده شده وجود دارد (هر تصویر اصلی دارای دو افزایش است).

¹⁶Embedding

¹⁷Projection Head

¹⁸Multilayer perceptron

¹⁹Feature Vector

²⁰Contrastive Loss

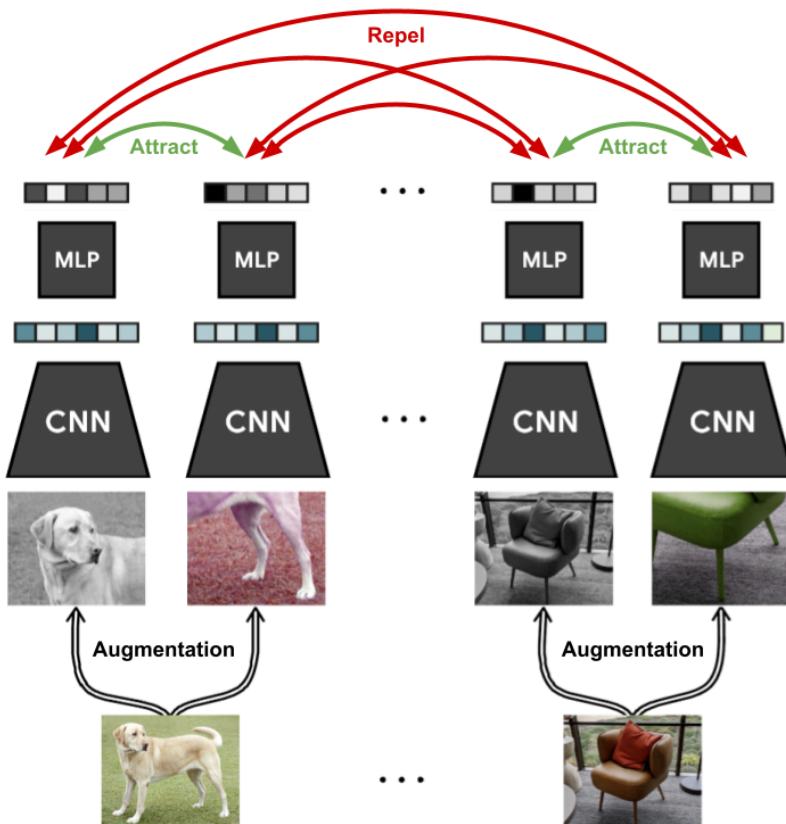
²¹Normalized Temperature-Scaled Cross-Entropy

²²Batch

شباهت کسینوسی ($\text{sim}(\mathbf{z}_i, \mathbf{z}_j)$) به صورت زیر محاسبه می‌شود:

$$\text{sim}(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i \cdot \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|}$$

از دست دادن NT-Xent اطمینان می‌دهد که جفت‌های مثبت (نمایه‌ای افزوده شده از یک تصویر) به مکان‌های مشابهی در فضای تعییه نگاشت می‌شوند، در حالی که جفت‌های منفی (نمایه‌ای تصاویر مختلف) به مکان‌های دور از هم نگاشت می‌شوند.

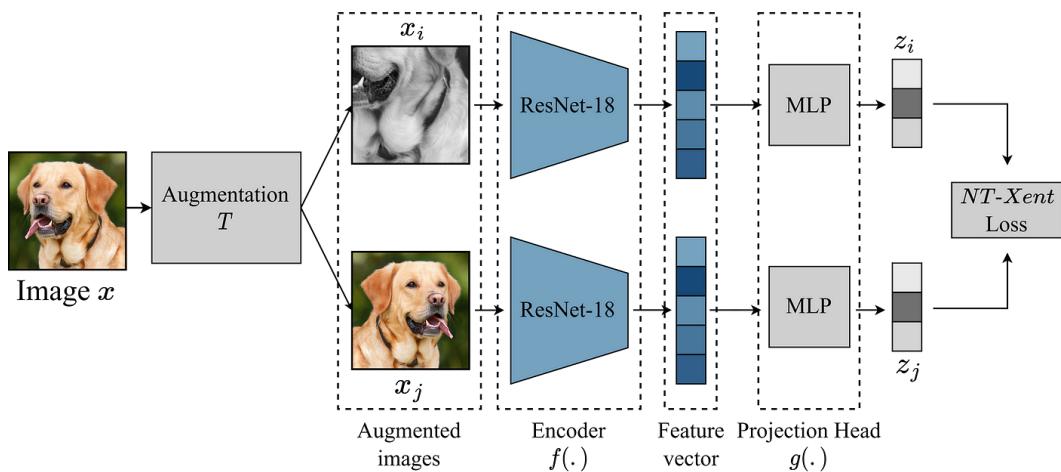


شکل ۳-۸: چارچوب ساده برای یادگیری تمایزی بازنمایی‌های بصری [۷]

- آموزش: در طول آموزش، SimCLR دو افزایش مختلف از هر تصویر در دسته تولید می‌کند. این جفت‌های افزوده شده از طریق رمزگذار و سرپیش‌بینی عبور می‌کنند تا تعییه‌ها به دست آیند. سپس از دست دادن تقابلی بر روی این تعییه‌ها محاسبه می‌شود. مدل با استفاده از گرادیان نزولی تصادفی آموزش داده می‌شود تا از دست دادن تمایزی به حداقل برسد.

- ارزیابی: پس از آموزش، سرپیش‌بینی دور انداخته می‌شود و رمزگذار برای وظایف پسین، مانند دسته‌بندی تصویر، استفاده می‌شود. بازنمایی‌های یادگرفته شده بسته به سناریو، تنظیم دقیق می‌شوند یا مستقیماً در

این وظایف استفاده می‌شوند.



[A]SimCLR ۹-۳: معماری

Barlow Twins ۴-۳

یک روش یادگیری خودنظراتی است که هدف آن یادگیری بازنمایی‌های معنادار از داده‌ها بدون نیاز به نمونه‌های برچسب‌گذاری شده است. این روش به‌طور خاص بر این تمرکز دارد که بازنمایی‌های یادگرفته شده توسط یک شبکه عصبی هم نسبت به تغییرات مختلف یک نقطه داده ثابت باشند و هم در میان نقاط داده مختلف تنوع داشته باشند.

مفهوم کلیدی Barlow Twins

- **کاهش افزونگی^{۲۳}:** ایده اصلی Barlow Twins این است که افزونگی در بازنمایی‌های یادگرفته شده را به حداقل برساند. به عبارت دیگر، این روش تلاش می‌کند تا اطمینان حاصل کند که ویژگی‌های مختلف داده ورودی توسط ابعاد مختلف بازنمایی یادگرفته شده به‌طور جداگانه و بدون همپوشانی ثبت شوند.
- **ماتریس همبستگی متقابل^{۲۴}:** از یک ماتریس همبستگی متقابل که بین بازنمایی‌های دو دیدگاه مختلف از یک ورودی محاسبه می‌شود، استفاده می‌کند. هدف این است که این ماتریس همبستگی متقابل تا حد ممکن به ماتریس همانی^{۲۵} نزدیک باشد. ماتریس همانی نشان می‌دهد که هر بعد از بازنمایی به‌طور مستقل از سایر ابعاد است و در نتیجه افزونگی به حداقل می‌رسد.
- **تابع هدف:** تابع زیانی که در Barlow Twins استفاده می‌شود به گونه‌ای طراحی شده است که هرگونه انحراف از ماتریس همانی در ماتریس همبستگی متقابل را مجازات کند. این کار شبکه را ترغیب می‌کند تا ویژگی‌هایی بدون همبستگی تولید کند، به‌طوری که هر بعد ویژگی از نظر آماری مستقل باشد.

تابع زیان Barlow Twins

تابع زیان در Barlow Twins [۹] به صورت زیر تعریف می‌شود:

$$L = \sum_i (1 - C_{ii})^{\gamma} + \lambda \sum_{i \neq j} C_{ij}^{\gamma}$$

که در آن:

- C ماتریس همبستگی متقابل محاسبه شده بین بازنمایی‌های دو دیدگاه مختلف است.

²³Redundancy Reduction

²⁴Cross-Correlation Matrix

²⁵Identity Matrix

• عناصر قطری ماتریس خودهمبستگی^{۲۶} را نشان می‌دهد.

• عناصر خارج از قطر هم‌بستگی متقابل را نشان می‌دهد.

• λ یک فاکتور وزنی است که اهمیت دو عبارت در تابع زیان را متعادل می‌کند.

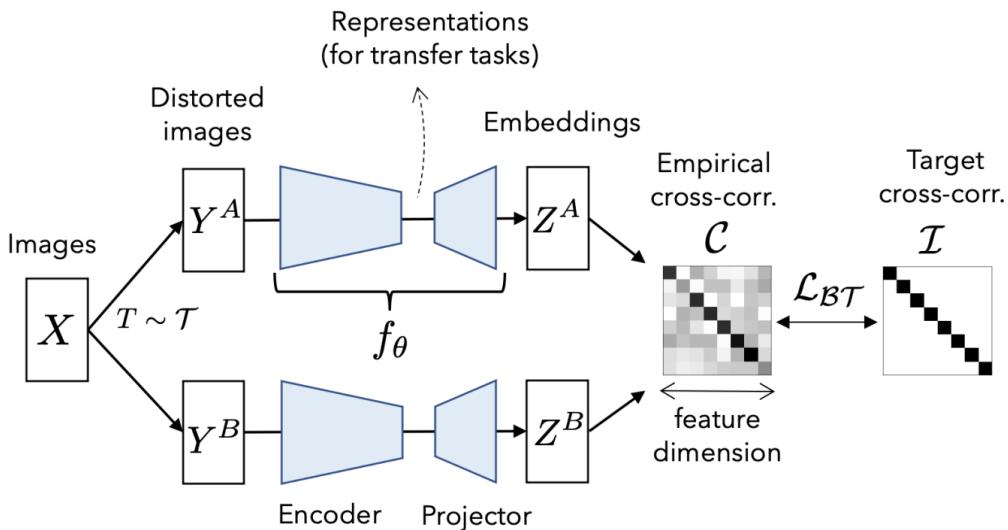
عبارت اول $C_{ii}^2 - 1$ اطمینان حاصل می‌کند که عناصر قطری ماتریس هم‌بستگی متقابل به ۱ نزدیک هستند، به این معنا که بازنمایی‌ها از تغییرات مختلف یک ورودی هم‌بستگی بالایی دارند.

عبارت دوم $\sum_{i \neq j} C_{ij}^2 \lambda$ عناصر خارج از قطر را مجازات می‌کند و آن‌ها را به سمت صفر می‌برد که به کاهش افزونگی در ویژگی‌های یادگرفته شده کمک می‌کند.

²⁶Auto-Correlation

Barlow Twins مراحل

- افزایش داده: مشابه Barlow Twins SimCLR با اعمال دو افزایش مختلف به یک تصویر ورودی شروع می‌شود و دو نمای مختلف ایجاد می‌کند.
- استخراج ویژگی: این دو نمای افزوده شده سپس از طریق همان شبکه عصبی عبور می‌کنند تا دو مجموعه بازنمایی ویژگی تولید شود.
- محاسبه همبستگی متقابل: بازنمایی‌ها نرمال شده و سپس برای محاسبه یک ماتریس همبستگی متقابل استفاده می‌شوند.
- محاسبه زیان: تابع زیان، همان‌طور که در بالا توضیح داده شد، برای کاهش افزونگی و تغییب به بازنمایی‌های متنوع محاسبه می‌شود.
- بهینه‌سازی: پارامترهای شبکه عصبی به روزرسانی می‌شوند تا زیان به حداقل برسد و از این طریق بازنمایی‌های مفید به صورت خودناظارتی یاد گرفته شوند.



[۹] شکل ۱۰-۳: Barlow Twins معماری

فصل چهارم

نتایج

۱-۴ مقدمه

در این فصل، معیارهای عملکرد مدل‌هایی که در طول این مطالعه آموزش داده شده‌اند را ارائه می‌دهیم. مدل‌ها بر اساس معیارهای خطای آموزش^۱ و خطای آزمایش^۲ ارزیابی می‌شوند که به ترتیب نشان‌دهنده عملکرد مدل بر روی داده‌های آموزشی و داده‌های آزمایش دیده‌نشده است. علاوه بر این، معیارهای دیگری مانند دقت^۳، صحت^۴، یادآوری^۵، و امتیاز F1^۶ نیز برای ارزیابی جامع‌تر عملکرد مدل‌ها محاسبه و تحلیل خواهند شد. نتایج را در میان معماری‌ها و روش‌های آموزشی مختلف، از جمله یادگیری با نظارت (ResNet50 و VGG16) و یادگیری خودناظارتی (Barlow Twins و SimCLR)، با تمرکز ویژه بر تأثیر استفاده از رمزگذارهای ثابت^۷ در مقابل رمزگذارهای آزاد در یادگیری خودناظارتی مقایسه می‌کنیم.

۲-۴ مدل‌های یادگیری با نظارت

۱-۲-۴ نتایج مدل ResNet50

نتایج ثابت داده‌های آزمایش در نمودار معیارهای دقت، صحت، یادآوری و امتیاز F1 در طول دوره‌ها^۸ به این دلیل است که ارزیابی مدل در انتهای داده‌های آزمایش صورت گرفته و در آموزش نقشی ندارد. این روش برای ارائه یک نقطه مرجع پایدار برای مقایسه عملکرد مدل در طول فرآیند آموزش انتخاب شد. همچنین دلیل اتمام دوره‌ها در دوره‌ی پانزدهم به علت وجود توقف زودهنگام^۹ است که برای صرفه‌جویی در زمان و محاسبات درنظر گرفته شده است.

¹Train Loss

²Test Loss

³Accuracy

⁴Precision

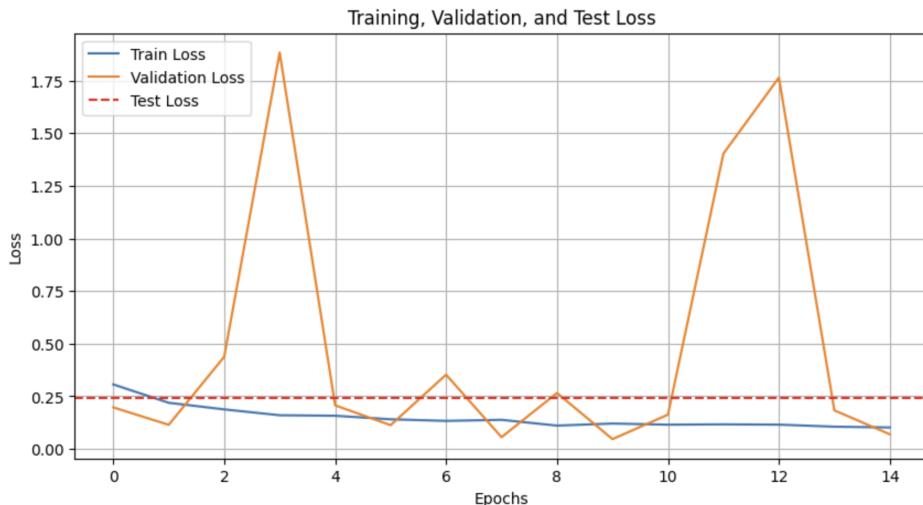
⁵Recall

⁶F1-Score

⁷Frozen

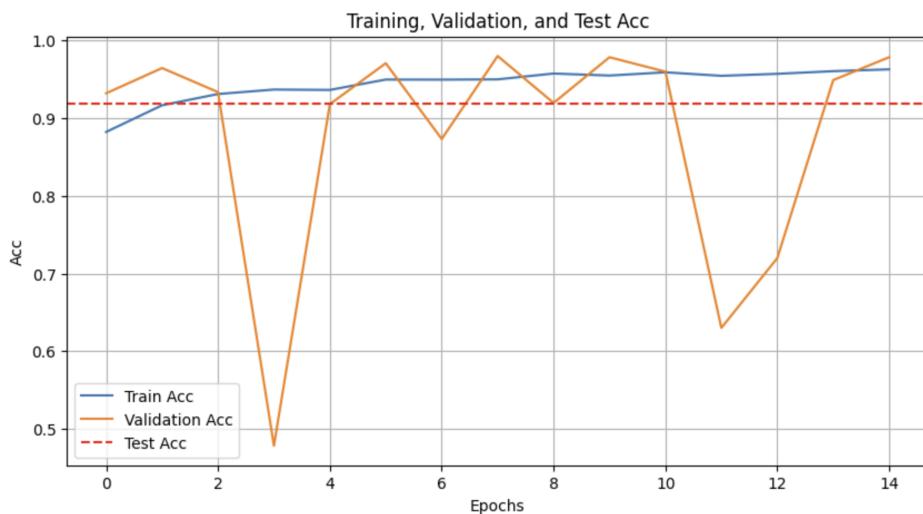
⁸Epochs

⁹Early Stopping



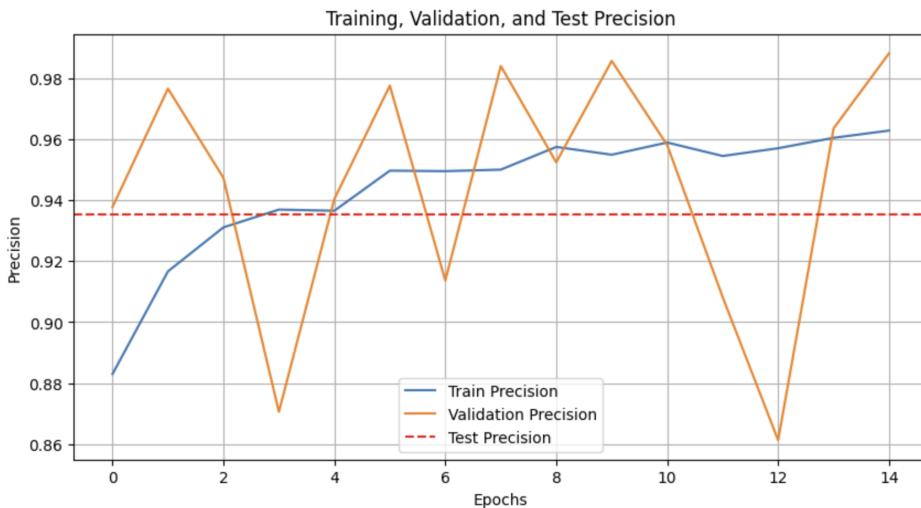
شکل ۱-۴: خطای آموزش، اعتبارسنجی و آزمایش

خطای آموزش به طور مداوم در طول زمان کاهش می‌یابد، که انتظار می‌رود زیرا مدل به مرور یاد می‌گیرد و بهتر با داده‌های آموزش منطبق می‌شود. با این حال، خطای اعتبارسنجی نوسانات نامنظمی را نشان می‌دهد که بار دیگر به مسئله بیش‌برازش اشاره دارد؛ جایی که مدل در داده‌های آموزشی عملکرد خوبی دارد اما در داده‌های اعتبارسنجی ضعیف عمل می‌کند.



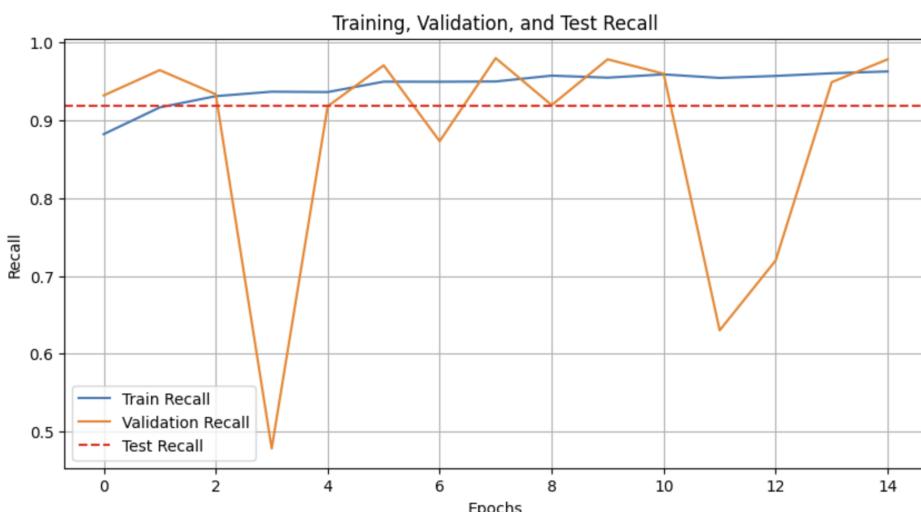
شکل ۲-۴: دقت آموزش و اعتبارسنجی و آزمایش

دقت آموزش به تدریج در طول دوره‌ها افزایش یافته و به بالای ۹۰٪ می‌رسد. با این حال، دقت اعتبارسنجی به طور قابل توجهی نوسان دارد که نشان‌دهنده احتمال بروز مشکل در بیش‌برازش است. در حالی که دقت آموزش بالا است، دقت اعتبارسنجی در برخی نقاط به شدت کاهش می‌یابد که نشان می‌دهد مدل ممکن است در تعمیم به داده‌های نادیده به خوبی عمل نکند.



شکل ۴-۳: صحت آموزش و اعتبارسنجی و آزمایش ResNet50

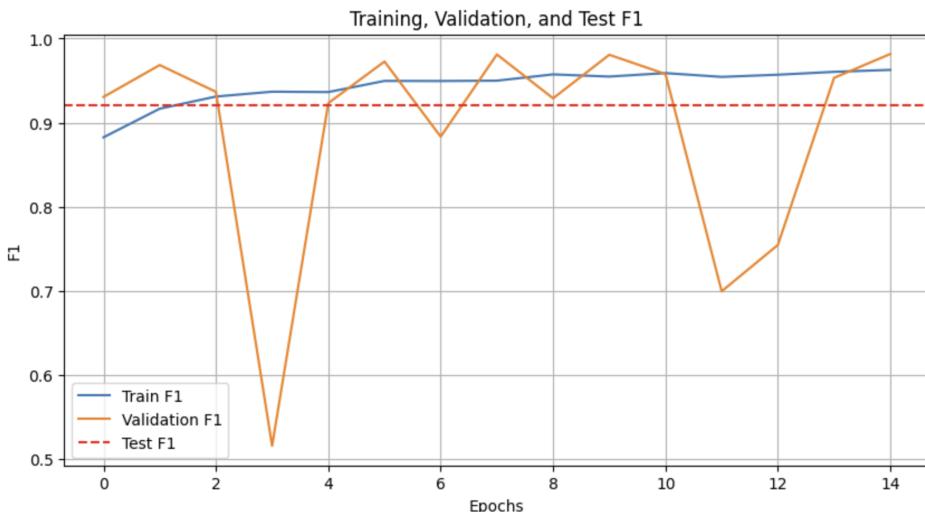
معیارهای صحت نشان می‌دهند که مدل در طول آموزش دقت بالایی را حفظ می‌کند، اما صحت در اعتبارسنجی به طور قابل توجهی نوسان دارد. این موضوع نشان می‌دهد که در حالی که مدل در پیش‌بینی‌های مثبت در طول آموزش دقیق است، در داده‌های اعتبارسنجی با مشکل روبرو می‌شود که ممکن است به دلیل بیش‌برازش باشد.



شکل ۴-۴: یادآوری آموزش و اعتبارسنجی و آزمایش ResNet50

معیار یادآوری نیز مشابه صحت در طول آموزش بهبود پایداری نشان می‌دهد. با این حال، یادآوری در اعتبارسنجی نوسانات قابل توجهی را نشان می‌دهد و در برخی دوره‌ها کاهش چشمگیری دارد. این موضوع نشان می‌دهد که توانایی مدل در شناسایی صحیح نمونه‌های مثبت در مجموعه اعتبارسنجی به‌طور قابل توجهی متغیر است.

امتیاز F1 که تعادلی بین صحت و یادآوری است، الگوی مشابهی با دقت را دنبال می‌کند. امتیاز F1 در



شکل ۴-۵: امتیاز F1 آموزش و اعتبارسنجی و آزمایش ResNet50

آموزش بالا و پایدار باقی می‌ماند، اما امتیاز F1 در اعتبارسنجی افت‌های قابل توجهی را تجربه می‌کند. این نوسان نشان‌دهنده ناپایداری مدل در حفظ تعادل بین صحت و یادآوری در بین کلاس‌های مختلف است.

نتیجه نهایی

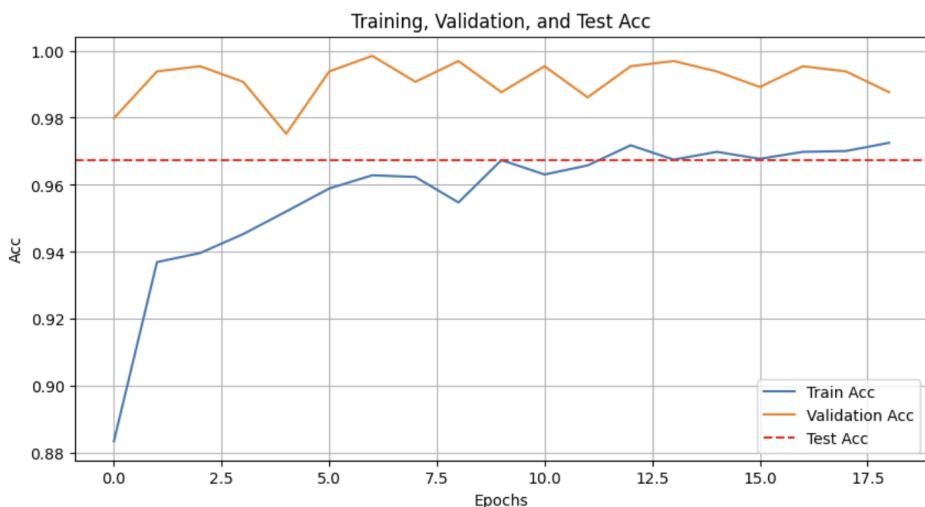
عملکرد بالا در تمامی معیارهای آزمایش از جمله دقت، صحت، یادآوری و امتیاز F1 نشان می‌دهد که مدل به خوبی آموزش داده شده و قادر به انجام پیش‌بینی‌های دقیق و قابل اعتماد است. نتایج نشان می‌دهند که مدل برای کاربردهای عملی که حفظ سطح بالایی از عملکرد در چندین معیار ارزیابی حیاتی است، مناسب است.

۴-۲-۴ نتایج مدل VGG16



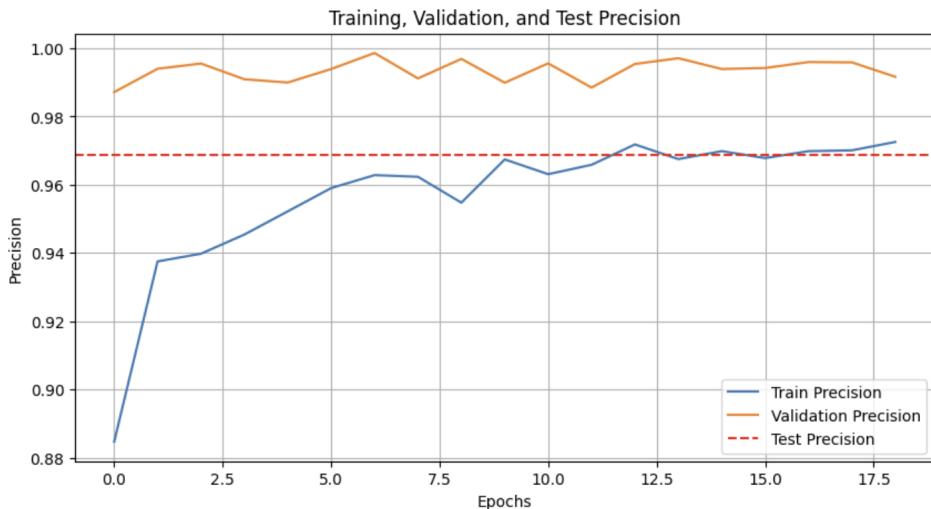
شکل ۴-۶: خطای آموزش، اعتبارسنجی و آزمایش VGG16

نمودار خطا نشان می‌دهد که خطای آموزش به طور پیوسته کاهش می‌یابد که نشان‌دهنده فرآیند یادگیری مدل است. با این حال، خطای اعتبارسنجی در چندین نقطه جهش‌هایی را نشان می‌دهد که بیشتر نشان‌دهنده احتمال بروز بیش‌بازش است. با وجود این، خطای آزمایش نسبتاً پایین و پایدار باقی می‌ماند که تأیید می‌کند مدل بر روی داده‌های جدید عملکرد خوبی دارد.



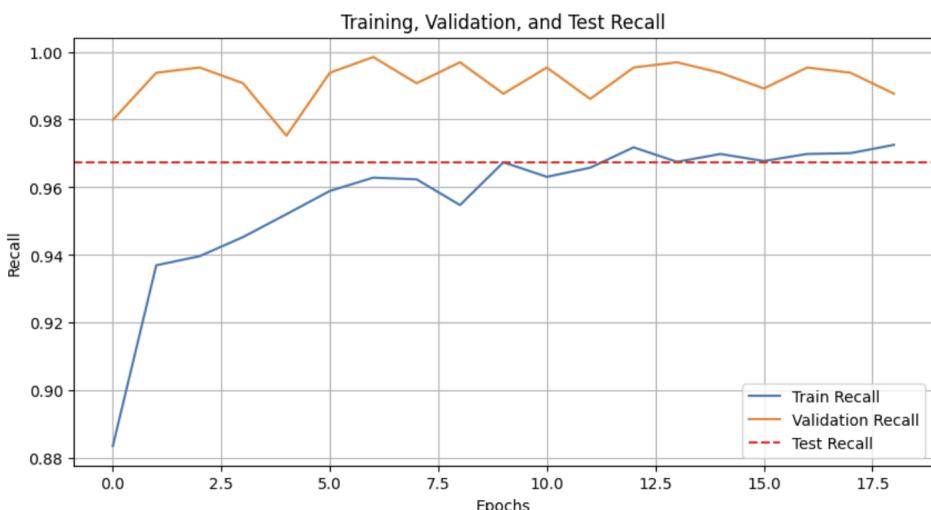
شکل ۷-۴: دقت آموزش و اعتبارسنجی و آزمایش VGG16

در نمودار دقت، دقت آموزش به طور مداوم بالا باقی می‌ماند و تنها نوسانات جزئی دارد، که نشان‌دهنده این است که مدل به طور مؤثری از داده‌های آموزشی یاد می‌گیرد. روند کلی نشان می‌دهد که مدل تثبیت می‌شود و دقت آزمایش، که با خط قرمز خط‌چین نشان داده شده است، بالای ۹۶٪ باقی می‌ماند که نشان‌دهنده عملکرد قوی بر روی داده‌های نادیده است.



شکل ۴-۸: صحت آموزش و اعتبارسنجی و آزمایش VGG16

نمودار صحت نشان می‌دهد که دقت مدل بر روی مجموعه آموزشی به مرور زمان بهبود می‌پابد و تثبیت می‌شود. صحت آزمایش بالا است که نتیجه مثبتی برای توانایی مدل در اجتناب از مثبت‌های کاذب است.



شکل ۴-۹: یادآوری آموزش و اعتبارسنجی و آزمایش VGG16

نمودار یادآوری عملکردی بالا و پایدار بر روی داده‌های آموزشی را نشان می‌دهد که مشابه با دقت و امتیاز F1 است. یادآوری در اعتبارسنجی نوساناتی را نشان می‌دهد، بهویژه با کاهش‌های تند در دوره‌های خاص، که نشان‌دهنده چالش‌های گاهوبیگاه مدل در شناسایی همه نمونه‌های مثبت مرتبط است. با این حال، یادآوری آزمایش، مانند سایر معیارها، بالای ۹۶٪ باقی می‌ماند که نشان می‌دهد مدل توانایی قوی در شناسایی صحیح نمونه‌های مثبت بر روی داده‌های جدید دارد.

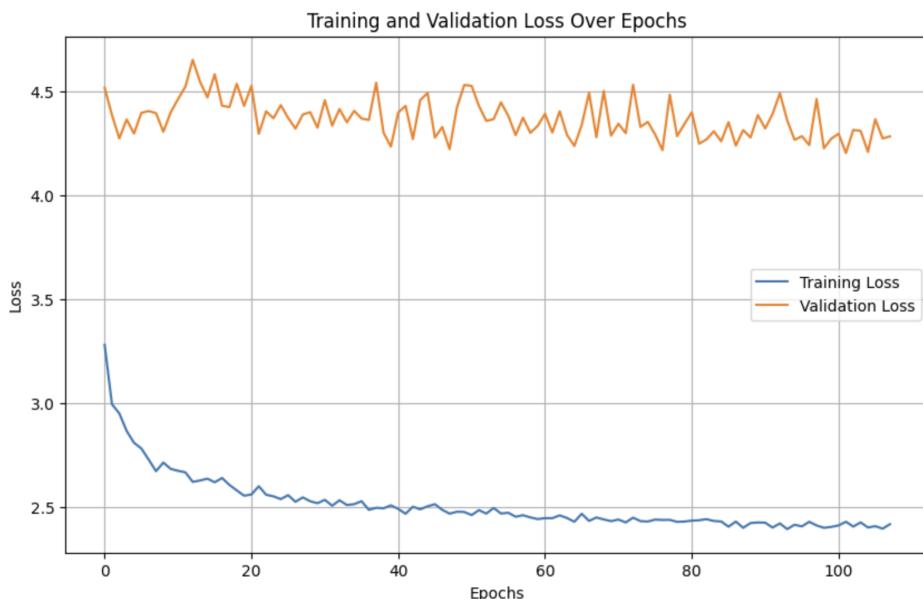


شکل ۴-۱۰: امتیاز F1 آموزش و اعتبارسنجی و آزمایش VGG16

نمودار امتیاز F1 الگوی مشابهی با نمودار دقت دارد، به طوری که امتیاز F1 آموزش بالا و پایدار باقی می‌ماند که نشان‌دهنده اثربخشی مدل در مدیریت هر دو معیار صحت و یادآوری است. امتیاز F1 آزمایش که همچنان بالای ۹۶٪ باقی می‌ماند، تأیید می‌کند که مدل تعادل خوبی بین صحت و یادآوری بر روی داده‌های نادیده حفظ می‌کند.

۳-۴ یادگیری خودناظارتی

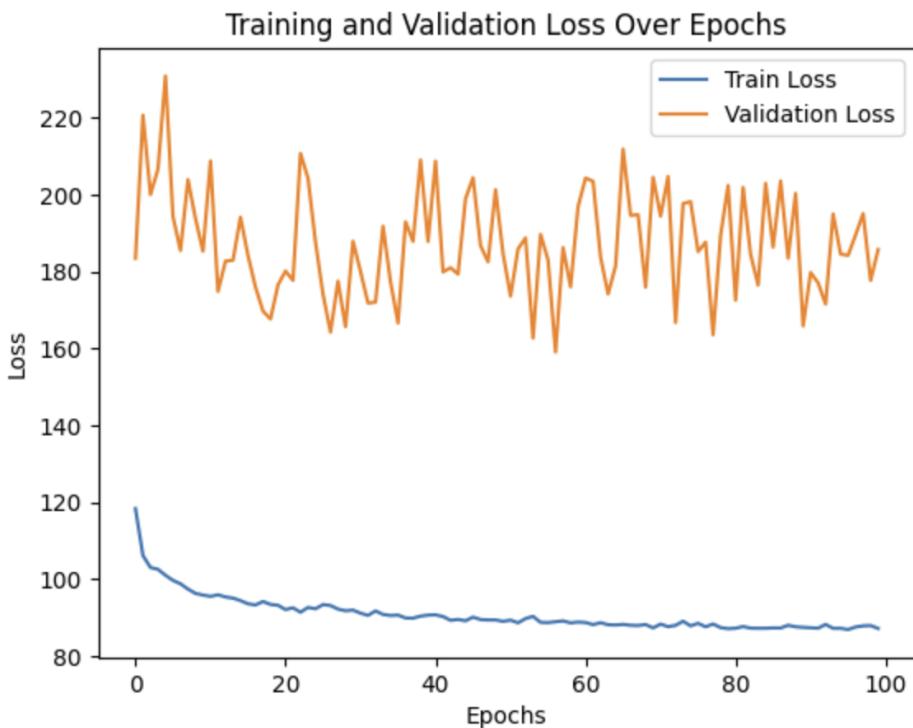
نتایج مدل SimCLR در آموزش



شکل ۱۱-۴: خطای آموزش در طول دوره‌ها در SimCLR

تحلیل خطای آموزش و اعتبارسنجی نشان می‌دهد که در حالی که مدل به طور مؤثری بر روی داده‌های آموزشی یاد می‌گیرد، توانایی آن در تعمیم به داده‌های اعتبارسنجی ممکن است محدود باشد. این ممکن است به دلیل بیشبرازش باشد، جایی که مدل بیش از حد به داده‌های آموزشی خاص می‌شود. اعمال تکنیک‌های منظم‌سازی یا تنظیم پارامترهای مدل می‌تواند به بهبود عملکرد آن بر روی داده‌های نادیده کمک کند.

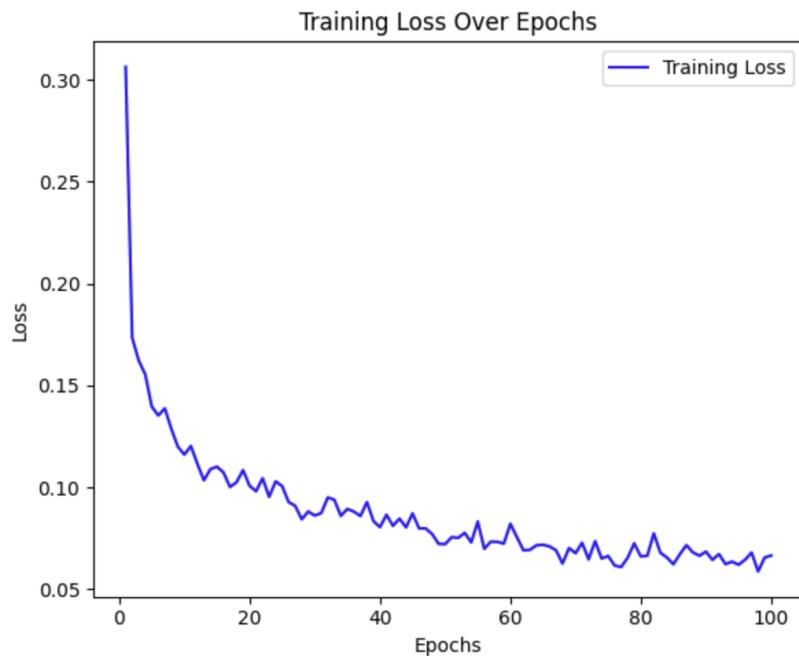
نتایج مدل Barlow Twins در آموزش



شکل ۱۲-۴: خطای آموزش در طول دوره‌ها در Barlow Twins

تحلیل خطای آموزش و اعتبارسنجی Barlow Twins نشان می‌دهد که مدل به‌طور مؤثر بر روی داده‌های آموزشی یاد می‌گیرد، اما در تعمیم دادن به داده‌های اعتبارسنجی دچار چالش‌هایی است، همان‌طور که از طریق نوسانات خطای اعتبارسنجی مشخص می‌شود. در مقایسه با SimCLR تمایل به نمایش نوسانات بیشتری در خطای اعتبارسنجی دارد که می‌تواند به تفاوت‌های موجود در توابع خطا نسبت داده شود. در حالی که SimCLR ممکن است پایداری و کارآیی بیشتری در آموزش ارائه دهد، Barlow Twins مزیت منحصر به‌فردی در یادگیری نمایه‌های ویژگی مستقل و غیر همبسته ارائه می‌دهد.

با توجه به اینکه رمزگذار آزاد یا ثابت است، در ادامه بررسی می‌کنیم که مدل خودناظارتی چگونه بر روی داده‌ها عمل خواهد کرد. این بررسی به ما اجازه می‌دهد تا تأثیر ثابت یا آزاد بودن رمزگذار بر عملکرد مدل را در فرآیند یادگیری و تعمیم‌دهی به داده‌های جدید مشاهده کنیم. در حالت ثابت، تنها لایه‌های نهایی مدل آموزش داده می‌شوند، در حالی که در حالت آزاد، تمامی لایه‌های مدل از جمله رمزگذار قابلیت تنظیم مجدد را دارند. این تفاوت می‌تواند تأثیر قابل توجهی بر روی دقت، خطا، و تعمیم‌پذیری مدل داشته باشد.



شکل ۱۳-۴: خطای آموزش در طول دوره‌ها در آزاد SimCLR

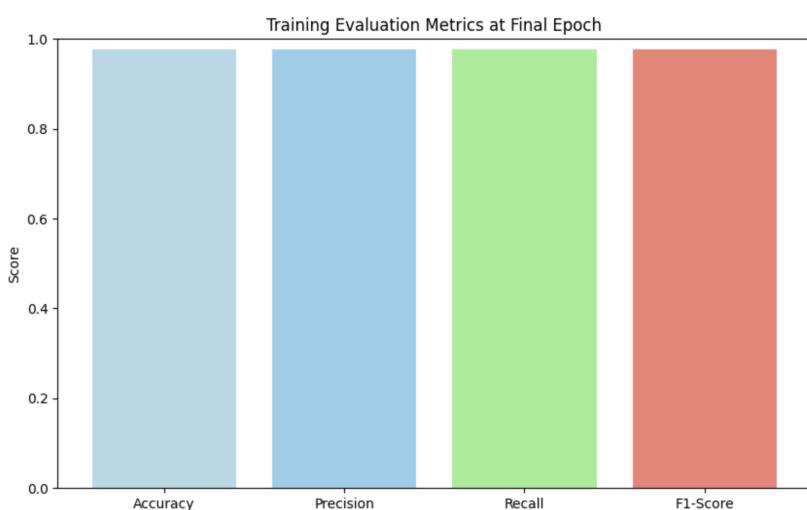
۱-۳-۴ نتایج مدل SimCLR با رمزگذار آزاد

منحنی خطا مکمل نمودار دقت است و کاهش شدید خطا در دوره‌های اولیه را نشان می‌دهد که با گذشت زمان به تدریج مسطح می‌شود. خطای نهایی بسیار کم، حدود 5.0° است که نشان می‌دهد پیش‌بینی‌های مدل بسیار نزدیک به برچسب‌های واقعی است. این نشانه مثبتی از همگرایی است و نشان می‌دهد که مدل الگوهای اساسی در داده‌ها را به خوبی آموخته است.



شکل ۱۴-۴: دقت آموزش در طول دوره‌ها مدل آزاد SimCLR

نمودار دقت نشان‌دهنده افزایش پایدار در طول فرآیند آموزش است که از حدود ۹۰٪ شروع شده و نهایتاً در حدود ۹۸٪ در دوره ۱۰۰ام تثبیت می‌شود. این روند نشان می‌دهد که مدل به‌طور مؤثری از داده‌های آموزشی یاد می‌گیرد. عدم وجود نوسانات قابل توجه در این منحنی نشان می‌دهد که مدل به‌طور پیوسته در حال بهبود است بدون اینکه با مشکلات جدی بیشبرازش یا کم برآذش مواجه شود.



شکل ۱۵-۴: معیارهای ارزیابی آموزش در دوره نهایی مدل آزاد SimCLR

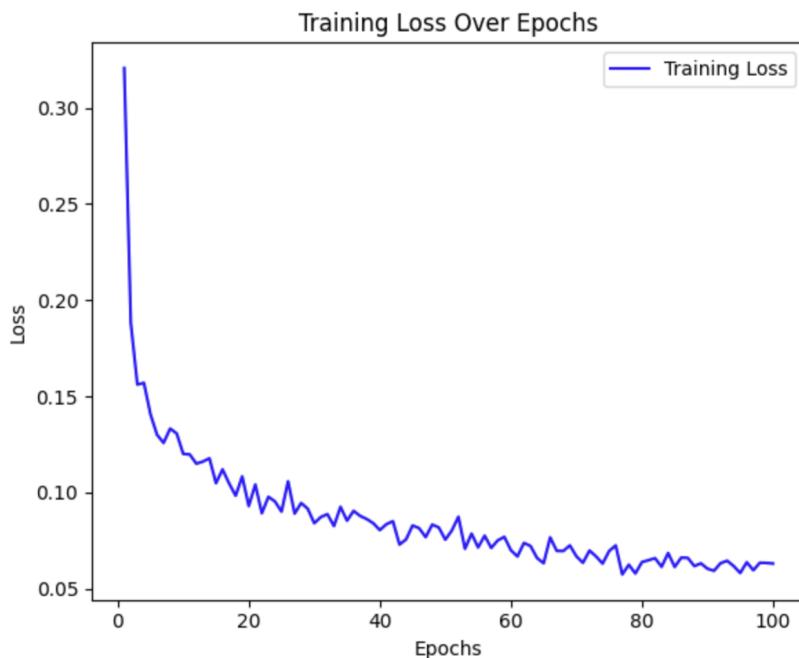
این نمودار میله‌ای نمای کلی از معیارهای ارزیابی نهایی آموزش برای SimCLR با رمزگذار ثابت آزاد را ارائه می‌دهد. دقت، صحت، یادآوری، و امتیاز F1 همگی نزدیک به ۱.۰ گزارش شده‌اند که نشان می‌دهد مدل در طول آموزش عملکرد بسیار خوبی داشته است. ثبات در این معیارها نشان می‌دهد که مدل نه تنها دقیق است بلکه دارای صحت بالا و یادآوری قوی نیز هست که نشان‌دهنده توانایی بالای مدل در پیش‌بینی صحیح برچسب‌های کلاس به‌طور مداوم است.

نتیجه نهایی

با رمزگذار آزاد عملکرد قوی‌ای در تمامی معیارها نشان داده است که اثربخشی آن را در این وظیفه نشان می‌دهد. با توجه به اینکه مدل از ResNet50 که بر روی ImageNet از پیش آموزش دیده است استفاده می‌کند، از نمایش‌های ویژگی قدرتمندی بهره می‌برد که به‌خوبی به این وظیفه تعمیم می‌یابد. دقت بالا، صحت، یادآوری، و امتیاز F1 در دوره نهایی نشان می‌دهند که مدل به‌خوبی کالیبره شده و قادر به انجام پیش‌بینی‌های قابل اعتماد است. منحنی‌های آموزشی صاف نیز این نتیجه‌گیری را پشتیبانی می‌کنند که مدل به‌طور مؤثری یاد گرفته است بدون اینکه بیش‌برازش قابل توجیه داشته باشد.

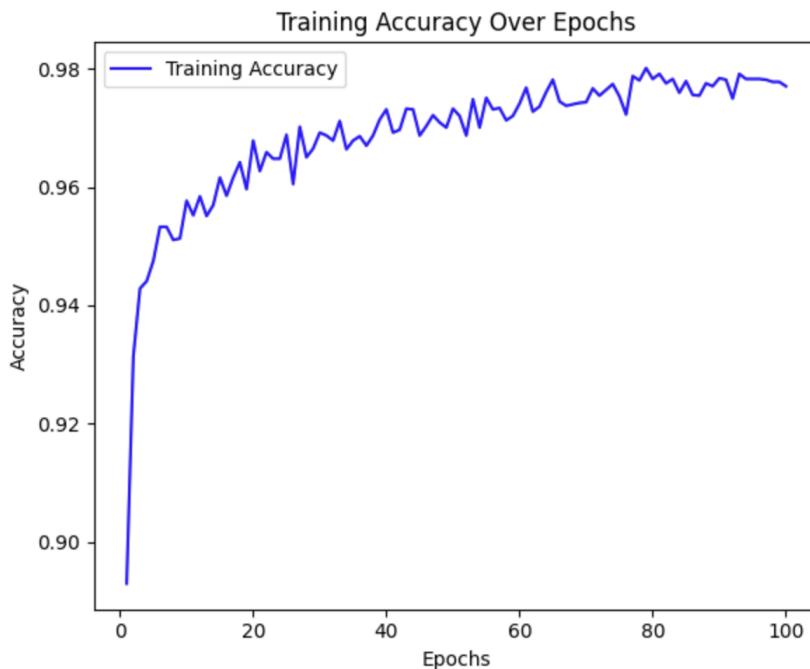
وقتی رمزگذار آزاد است، کل شبکه، شامل لایه‌های از پیش آموزش دیده ResNet50، بر روی وظیفه جدید تنظیم دقیق می‌شود. این اجزا می‌دهد که مدل به‌طور مؤثرتری به خصوصیات مجموعه داده‌ها تطبیق یابد. با این حال، این فرآیند از نظر محاسباتی پرهزینه‌تر است و زمان بیشتری می‌برد زیرا گرادیان‌ها باید در تمامی لایه‌ها محاسبه و به‌روزرسانی شوند.

۲-۳-۴ نتایج مدل SimCLR با رمزگذار ثابت



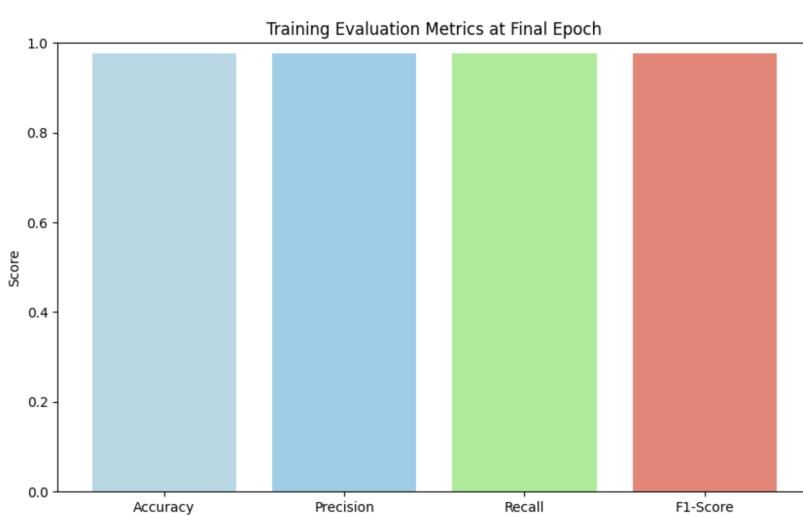
شکل ۱۶-۴: خطای آموزش در طول دوره‌ها در SimCLR ثابت

نشان می‌دهد که خطای آموزش به مرور زمان به طور پیوسته کاهش می‌یابد. خطای در ابتدا نسبتاً بالا شروع می‌شود اما به سرعت در طی چند دوره اول کاهش می‌یابد که نشان‌دهنده یادگیری مؤثر است. خطای مدل به کاهش ادامه می‌دهد، هرچند با نرخ کندتر، همان‌طور که آموزش پیش می‌رود. این نشان می‌دهد که حتی با رمزگذار ثابت، لایه نهایی مدل به طور مؤثری خطا را به حداقل می‌رساند و منجر به پیش‌بینی‌های دقیق می‌شود.



شکل ۱۷-۴: دقت آموزش در طول دوره‌ها مدل SimCLR ثابت

نشان می‌دهد که چگونه دقت آموزش در طول دوره‌ها تکامل یافته است. نمودار نشان‌دهنده افزایش پایدار دقت در دوره‌های اولیه است، با مسطح شدن منحنی وقتی که مدل به حداقل دقت خود نزدیک می‌شود. مدل به سرعت به سطوح بالای دقت (بیش از ۹۵٪) در اولین ۲۰ دوره می‌رسد، که نشان‌دهنده یادگیری کارآمد است. با این حال، از آنجا که رمزگذار ثابت است، توانایی مدل برای بهبود بیشتر به لایه نهایی محدود می‌شود و منجر به همگرایی کمی کندر در مقایسه با مدل‌های با رمزگذار آزاد می‌شود.



شکل ۱۸-۴: معیارهای ارزیابی آموزش در دوره نهایی در SimCLR ثابت

عملکرد نهایی مدل SimCLR با رمزگذار ثابت را پس از آموزش نشان می‌دهد. معیارهای دقت، صحت، یادآوری

و امتیاز F1 همگی بسیار بالا و نزدیک به ۱۰٪ هستند، که نشان می‌دهد مدل در مجموعه داده‌های آموزشی عملکرد فوق العاده‌ای داشته است. این نتایج نشان می‌دهند که حتی با رمزگذار ثابت و با ترکیب یک لایه نهایی به خوبی تنظیم شده، می‌توان به عملکرد عالی در یک چارچوب یادگیری خودناظارتی دست یافت.

کارآیی

یکی از مزایای اصلی استفاده از رمزگذار ثابت در این تنظیمات، کاهش زمان آموزش است. رمزگذار ثابت تعداد پارامترهایی که باید آموزش داده شوند را محدود می‌کند و این به طور قابل توجهی پیچیدگی محاسباتی را کاهش می‌دهد. در مقایسه با رمزگذار آزاد، فرآیند آموزش این مدل تقریباً در نصف زمان به پایان رسید که آن را بسیار کارآمدتر می‌کند.

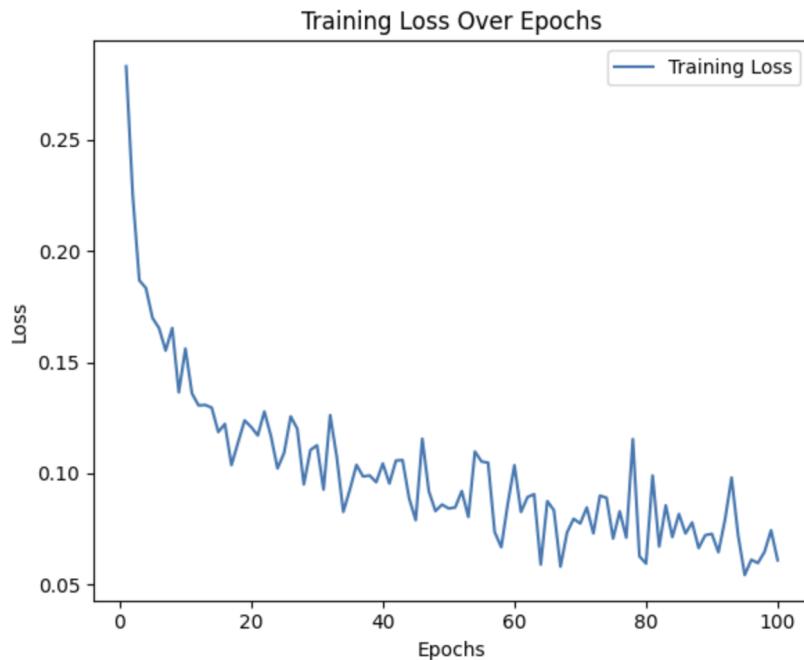
عملکرد

اگرچه ثابت نگهداشت رمزگذار سرعت فرآیند آموزش را افزایش می‌دهد، ممکن است توانایی مدل در تنظیم دقیق استخراج ویژگی‌ها از داده‌ها را محدود کند. با این حال، در این مورد، نتایج نشان می‌دهند که مدل همچنان به عملکرد بالایی در تمام معیارها دست می‌یابد و آن را به گزینه‌ای قابل قبول برای موقعیت‌هایی که منابع محاسباتی محدود یا نیاز به تکرار سریع وجود دارد تبدیل می‌کند.

نتیجه نهایی

به طور کلی، مدل SimCLR با رمزگذار ثابت عملکرد عالی‌ای را در تمامی معیارهای کلیدی نشان می‌دهد. این مدل به‌ویژه در سناریوهایی که سرعت آموزش در اولویت است، مفید است زیرا زمان آموزش را تقریباً به نصف در مقایسه با مدل‌های با رمزگذار آزاد کاهش می‌دهد. با وجود محدودیت‌های احتمالی در قابلیت‌های تنظیم دقیق، دقت، صحت، یادآوری و امتیاز F1 مدل همچنان بسیار بالا باقی می‌ماند که نشان می‌دهد ثابت نگهداشت رمزگذار به طور قابل توجهی اثربخشی مدل را کاهش نمی‌دهد.

۳-۳-۴ نتایج مدل Barlow Twins با رمزگذار آزاد



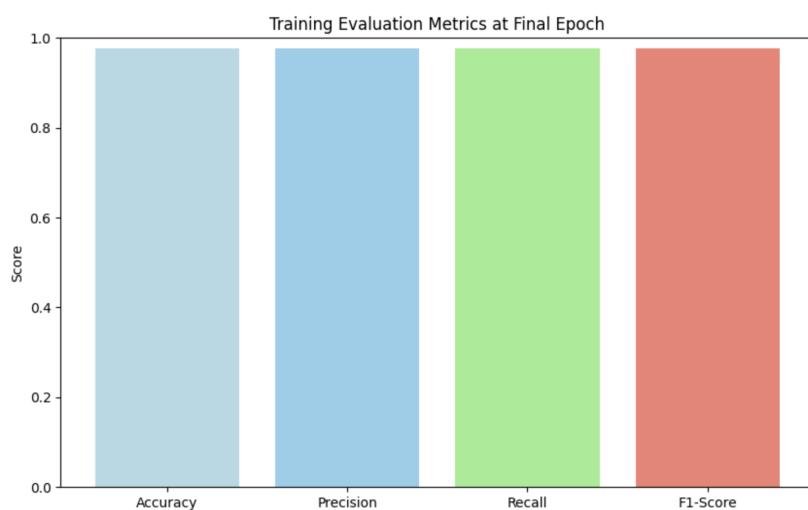
شکل ۱۹-۴: خطای آموزش در طول دوره‌ها در Barlow Twins آزاد

خطای آموزش از مقدار بالاتری شروع می‌شود و به طور مداوم در طول دوره‌ها کاهش می‌یابد و در نهایت در مقدار پایینی نزدیک به 5.0° تثبیت می‌شود. این نشان می‌دهد که پیش‌بینی‌های مدل با پیشرفت آموزش به طور فرایندهای دقیق می‌شوند. رمزگذار آزاد به یادگیری مؤثر کمک می‌کند، همان‌طور که کاهش مداوم خطا نشان می‌دهد. مدل به طور موفقیت‌آمیز تفاوت بین پیش‌بینی‌ها و مقادیر واقعی را به حداقل می‌رساند که نشان‌دهنده موفقیت‌آمیز بودن آموزش است.



شکل ۲۰-۴: دقت آموزش در طول دوره‌ها مدل آزاد Barlow Twins

دقت آموزش به طور تدریجی در طول دوره‌ها بهبود می‌یابد و تا پایان آموزش به نزدیک ۹۸٪ می‌رسد. منحنی دقت نسبتاً هموار است و نوسانات جزئی آن با پیشرفت آموزش ثبت می‌شود. فرآیند آموزش پایدار است و مدل به طور مداوم با یادگیری از داده‌ها بهبود می‌یابد. رمزگذار آزاد امکان تنظیم دقیق‌تر و عمیق‌تر را فراهم می‌کند، که منجر به دقت نهایی بالاتری نسبت به رمزگذار ثابت می‌شود.



شکل ۲۱-۴: معیارهای ارزیابی آموزش در دوره نهایی در آزاد Barlow Twins

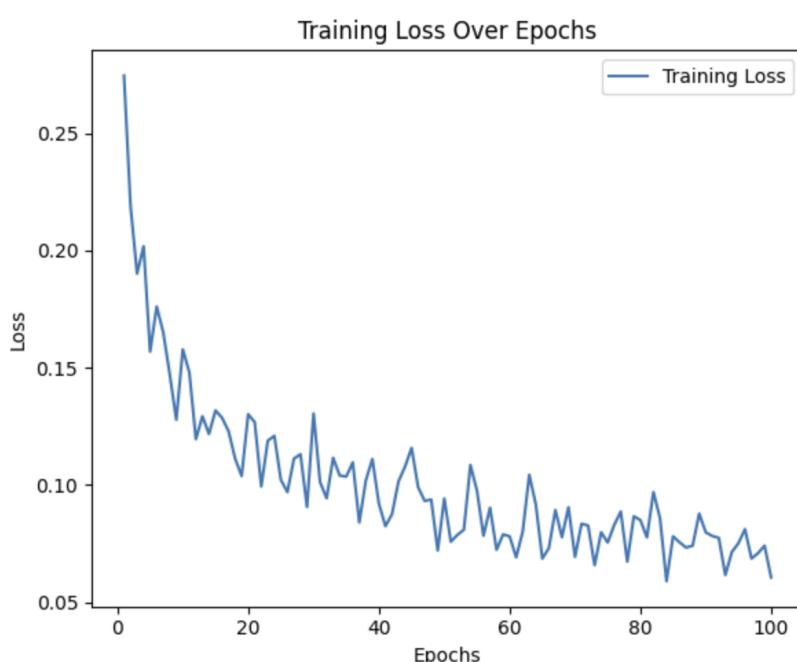
معیارهای نهایی آموزش برای دقت، صحت، فراخوانی و F1-Score همگی بالاتر از ۹۸٪ هستند. این نشان‌دهنده سطح عملکرد بسیار بالایی در تمام معیارهای ارزیابی است. مدل بسیار دقیق و سازگار در پیش‌بینی‌های خود

عمل می‌کند و این نشان می‌دهد که تنظیم دقیق رمزگذار به همراه بقیه شبکه به‌طور قابل توجهی به بهبود عملکرد کمک کرده است.

نتیجه نهایی

به طور کلی، مدل Barlow Twins با رمزگذار آزاد عملکرد برتری از نظر دقت، صحت، فراخوانی و F1-Score نشان می‌دهد و آن را به یک مدل قوی‌تر برای وظیفه مورد نظر تبدیل می‌کند. با این حال، این به هزینه افزایش منابع محاسباتی و زمان مورد نیاز برای آموزش به دست می‌آید.

۴-۳-۴ نتایج مدل Barlow Twins با رمزگذار ثابت



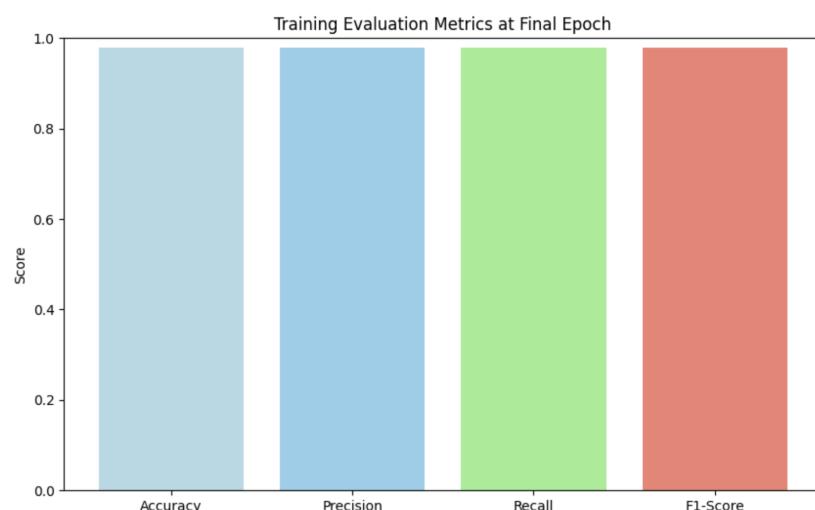
شکل ۲۲-۴: خطای آموزش در طول دوره‌ها در Barlow Twins ثابت

خطای آموزش با پیشرفت آموزش به‌طور مداوم کاهش می‌یابد و در دوره‌های آخر تنها نوسانات جزئی مشاهده می‌شود. روند کاهشی خطای آموزش نشان می‌دهد که مدل به‌طور پیوسته عملکرد خود را با به حداقل رساندن خطای آموزش می‌بخشد. نوسانات جزئی می‌تواند به طبیعت تصادفی فرآیند آموزش نسبت داده شود، اما به اندازه‌ای نیست که نشان‌دهنده مشکلات عمده‌ای باشد.



شکل ۲۳-۴: دقت آموزش در طول دوره‌ها مدل Barlow Twins ثابت

دقت آموزش به طور مداوم در طول دوره‌ها افزایش می‌یابد و در نهایت به حدود ۹۸٪ تثبیت می‌شود. این افزایش مداوم در دقت نشان می‌دهد که مدل به طور مؤثر از داده‌های آموزشی یاد می‌گیرد. عدم وجود نوسانات قابل توجه نشان‌دهنده یک فرآیند آموزش پایدار است که مشخصه مدل‌هایی است که به خوبی تنظیم شده‌اند.



شکل ۲۴-۴: معیارهای ارزیابی آموزش در دوره نهایی در Barlow Twins ثابت

معیارهای نهایی برای دقت، صحت، فراخوانی و F1-Score همگی بسیار بالا و در حدود ۹۹-۹۸٪ هستند. این نشان می‌دهد که مدل در وظایف طبقه‌بندی بسیار خوب عمل می‌کند. مقادیر بالای این معیارها نشان می‌دهد که مدل Barlow Twins، حتی با رمزگذار ثابت، قادر به یادگیری مؤثر از مجموعه داده است. ثبات در تمامی این

معیارها حاکی از آن است که مدل در پیش‌بینی‌های خود متعادل عمل می‌کند و بدون ترجیح دادن صحت به فراخوانی یا بر عکس، عملکردی متوازن دارد.

مقایسه با رمزگذار آزاد

• **سرعت آموزش:** با رمزگذار ثابت، تنها لایه‌های نهایی شبکه آموزش داده می‌شوند، که به طور قابل توجهی بار محاسباتی را کاهش می‌دهد. این امر منجر به یک فرآیند آموزشی می‌شود که تقریباً دو برابر سریع‌تر از زمانی است که رمزگذار آزاد باشد. این مزیت سرعت بهویژه در هنگام کار با مجموعه داده‌های بزرگ یا در مواقعي که منابع محاسباتي محدود هستند، بسیار مفید است.

• **کارآیی:** با وجود زمان آموزش سریع‌تر، مدل با رمزگذار ثابت همچنان به عملکرد بالایی دست می‌یابد، که نشان می‌دهد ثابت نگهداشت رمزگذار به طور قابل توجهی توانایی مدل در یادگیری از داده‌ها را به خطر نمی‌اندازد. این رویکرد را به یک روش بسیار کارآمد تبدیل می‌کند، بهویژه در سناریوهایی که آموزش سریع ضروری است.

نتیجه نهایی

مدل Barlow Twins با رمزگذار ثابت، کارآیی و عملکرد بالایی را نشان می‌دهد و به امتیازهای نزدیک به کامل در دقت، صحت، فراخوانی و F1-Score دست می‌یابد. فرآیند آموزش پایدار است و کاهش زمان آموزش این رویکرد را بهویژه برای کاربردهای عملی بسیار سودمند می‌سازد. عملکرد بالا، همراه با کارآیی محاسباتی، این تنظیم را برای سناریوهایی که زمان و منابع محدود هستند، ایده‌آل می‌کند.

فصل پنجم

جمع‌بندی و نتیجه‌گیری

۱-۵ مدل‌های بررسی شده

این گزارش به بررسی شش روشی شرکت مختلف آموزشی، شامل روش‌های یادگیری نظارت شده و یادگیری خودناظاری (SSL)، با استفاده از دو معماری اصلی ResNet50 و VGG16 پرداخته است. مدل‌های ارزیابی شده شامل:

۱. یادگیری نظارت شده با ResNet50

۲. یادگیری نظارت شده با VGG16

۳. SimCLR با رمزگذار ثابت (با معماری ResNet50)

۴. SimCLR با رمزگذار غیرثابت (با معماری ResNet50)

۵. Barlow Twins با رمزگذار ثابت (با معماری ResNet50)

۶. Barlow Twins با رمزگذار غیرثابت (با معماری ResNet50)

۲-۵ تحلیل عملکرد هر مدل

۱-۲-۵ مدل‌های یادگیری نظارت شده

مزایا:

• VGG16: دقیق بالاتر در معیارهای مختلف مانند امتیاز F1، صحت و یادآوری با نوسانات کمتر و ثبات بیشتر.

• ResNet50: ارائه دقیق بالا و تعمیم‌پذیری مناسب در کنار معماری پیچیده‌تر مناسب در مسائل خاص مفید.

معایب:

• VGG16: معماری ساده‌تر و عملکرد ضعیفتر در مسائل پیچیده نسبت به ResNet50.

• ResNet50: نوسانات بیشتر در برخی معیارها مانند امتیاز F1 و دقیق، همراه با پیچیدگی بالاتر به همراه هزینه محاسباتی بالاتر.

نتیجه‌گیری: با توجه به مزایا و معایب، VGG16 در این مقایسه عملکرد بهتری نشان داده است، زیرا دقیق و ثبات بیشتری در معیارهای کلیدی مانند امتیاز F1، صحت و یادآوری داشته و نوسانات کمتری در طول دوره‌های مختلف نشان داده است.

۲-۲-۵ مدل‌های SimCLR

- رمزگذار ثابت: زمان آموزش به طور قابل توجهی کاهش یافت به دلیل تعداد محدودی از پارامترهای به روزرسانی شده. با این حال، این امر با کاهش اندازی در دقت و افزایش خطای نسبت به نسخه غیرثابت همراه بود. مدل عملکرد خوبی داشت، اما از مزایای تنظیم دقیق کمتر برخوردار بود. فرآیند آموزش طولانی‌تر بود اما منجر به مدل دقیق‌تر و بهتر تنظیم‌شده‌ای شد.

مزایا:

- رمزگذار ثابت: آموزش سریع‌تر، هزینه محاسباتی کمتر.

- رمزگذار غیرثابت: دقت بالاتر، تنظیم دقیق‌تر.

معایب:

- رمزگذار ثابت: دقت کمتر نسبت به نسخه غیرثابت.

- رمزگذار غیرثابت: زمان آموزش طولانی‌تر، هزینه محاسباتی بیشتر.

نتیجه‌گیری: در حالی که نسخه غیرثابت SimCLR عملکرد بهتری ارائه داد، نسخه ثابت آن یک گزینه سریع‌تر و کارآمدتر از نظر محاسباتی را فراهم کرد. بسته به نیازهای خاص، هر یک از این مدل‌ها می‌تواند مزیت‌های خود را داشته باشد.

۳-۲-۵ مدل‌های Barlow Twins

- رمزگذار ثابت: مانند Barlow Twins، SimCLR با رمزگذار ثابت به دلیل به روزرسانی کمتر پارامترها سریع‌تر آموزش داده شد. دقت بالا بود، اما مدل نوسانات بیشتری در خطای اعتبارسنجی نشان داد که به معنای پایداری کمتر است.

- رمزگذار غیرثابت: عملکرد بهتری از نظر دقت، صحت، فراخوانی و F1-Score نسبت به نسخه ثابت داشت. آموزش پایدارتر بود و خطای مداوم کاهش یافت. با این حال، این نیاز به زمان آموزش طولانی‌تری داشت.

مزایا:

- رمزگذار ثابت: آموزش سریع‌تر، عملکرد خوب با منابع محاسباتی کمتر.

- رمزگذار غیرثابت: دقت بالاتر، فرآیند آموزش پایدار‌تر.

معایب:

- رمزگذار ثابت: نوسانات بیشتر در خطای اعتبارسنجی، دقت کمتر.

- رمزگذار غیرثابت: زمان آموزش طولانی‌تر، نیاز به منابع محاسباتی بیشتر.

نتیجه‌گیری: مشابه Barlow Twins، SimCLR با رمزگذار غیرثابت عملکرد کلی بهتری ارائه داد، در حالی که نسخه ثابت سریع‌تر و کم‌هزینه‌تر بود.

۳-۵ مقایسه دقت مدل‌ها با یکدیگر

مدل	دقت نهایی
یادگیری ناظارت‌شده ResNet50	۹۸.۰
یادگیری ناظارت‌شده VGG16	۹۹.۰
با رمزگذار ثابت SimCLR	۹۵.۰
با رمزگذار غیرثابت SimCLR	۹۸.۰
با رمزگذار ثابت Barlow Twins	۹۶.۰
با رمزگذار غیرثابت Barlow Twins	۹۸.۰

جدول ۱-۵: مقایسه دقت نهایی مدل‌ها

۴-۵ مقایسه زمانی مدل‌ها با یکدیگر

این مدل‌ها روی کولب با گرافیک پردازشی GPU T4 اجرا شده است و زمان‌ها با این معیار محاسبه شده‌اند.

مدل	توقف زودهنگام	تعداد دورها	زمان آموزش (ساعت)	زمان استنتاج (ثانیه)
یادگیری ناظارت‌شده ResNet50	فعال	۱۴	۴	۰.۵۲۵
یادگیری ناظارت‌شده VGG16	فعال	۱۸	۵	۰.۴۷۶
با رمزگذار ثابت SimCLR	غیرفعال	۱۰۰	۱۰:۳۰	۰.۴۲۸
با رمزگذار آزاد SimCLR	غیرفعال	۱۰۰	۱۲	۰.۴۲۸
با رمزگذار ثابت Barlow Twins	غیرفعال	۱۰۰	۱۰	۰.۳۸
با رمزگذار آزاد Barlow Twins	غیرفعال	۱۰۰	۱۲	۰.۳۸

جدول ۲-۵: مقایسه زمان‌های آموزش و استنتاج مدل‌ها به همراه توقف زودهنگام و تعداد دورها

۵-۵ مقایسه مدل‌های یادگیری نظارت شده با مدل‌های یادگیری خودناظارتی

در مقایسه مدل‌های یادگیری نظارت شده (ResNet50 و VGG16) با مدل‌های یادگیری خودناظارتی (SSL)، چندین نکته مهم به دست می‌آید:

۱-۵-۵ دقیق و عملکرد

- **مدل‌های یادگیری نظارت شده (VGG16، ResNet50):** این مدل‌ها به طور مداوم به دقیق و عملکرد بالایی در معیارهای مختلف (دقیق، صحت، فراخوانی، F1-Score) دست یافتند. به ویژه، VGG16 ثبات و توانایی تعمیم‌پذیری برتری نشان داد و از ResNet50 عملکرد بهتری داشت.

- **مدل‌های خودناظارتی (Barlow Twins، SimCLR):** دقیق و عملکرد مدل‌های خودناظارتی به طور کلی بالا بود، اگرچه در برخی موارد کمی کمتر از مدل‌های یادگیری نظارت شده بود. با این حال، این تفاوت قابل توجه نبود، به ویژه زمانی که رمزگذار غیرثابت و مدل به خوبی تنظیم دقیق شد.

۲-۵-۵ ثبات آموزش و کاهش خطأ

- **مدل‌های یادگیری نظارت شده:** هر دو مدل ResNet50 و VGG16 فرآیندهای آموزشی پابداری را نشان دادند و نوسانات کمی در خطای آموزش و اعتبارسنجی داشتند، که نشان‌دهنده یادگیری و تعمیم‌پذیری قوی است.

- **مدل‌های خودناظارتی:** ثبات آموزشی مدل‌های خودناظارتی بستگی به ثابت یا آزاد بودن رمزگذار داشت. نسخه‌های غیرثابت هر دو مدل Barlow Twins و SimCLR آموزش پایدارتر و نوسانات کمتری در کاهش خطأ داشتند.

۳-۵-۵ زمان آموزش و هزینه محاسباتی

- **مدل‌های یادگیری نظارت شده:** این مدل‌ها به طور کلی به منابع محاسباتی بیشتر و زمان آموزش طولانی‌تر نیاز دارند، به دلیل نیاز به داده‌های برچسب‌گذاری شده و معماهای پیچیده‌تر.

- **مدل‌های خودناظارتی:** مدل‌های خودناظارتی، به ویژه آن‌هایی که رمزگذار ثابت داشتند، کاهش قابل توجهی در زمان آموزش و هزینه محاسباتی ارائه دادند. این به دلیل بهروزرسانی پارامترهای کمتر بود که آموزش آن‌ها را سریع‌تر و کم‌هزینه‌تر کرد.

۴-۵-۵ هزینه برچسب‌گذاری و کاربرد پذیری

• مدل‌های یادگیری نظارت شده: در حالی که این مدل‌ها دقیق‌تر کمی بالاتری دارند، نیاز به مجموعه‌های داده برچسب‌گذاری شده گسترشده‌ای دارند که می‌تواند هزینه‌بر و زمان‌بر باشد. در شرایطی که برچسب‌ها کم‌یا در دسترس نیستند، مدل‌های یادگیری نظارت شده ممکن است قابل استفاده نباشند.

• مدل‌های خودناظارتی: مدل‌های خودناظارتی در شرایطی که داده‌های برچسب‌گذاری شده محدود یا غیرممکن است، بسیار کاربردی هستند. با وجود اینکه گاهی کمی کمتر دقیق هستند، آن‌ها یک جایگزین عملی و مقرون به صرفه ارائه می‌دهند. آن‌ها نیاز به داده‌های برچسب‌گذاری شده را با استفاده از مقادیر زیادی داده‌های بدون برچسب کاهش می‌دهند، که آن‌ها را در کاربردهای واقعی که برچسب‌گذاری دشوار یا بسیار گران است، بسیار ارزشمند می‌کند.

۴-۶ نتیجه‌گیری نهایی: ارزش مدل‌های خودناظارتی

با توجه به زمان و هزینه مرتبط با برچسب‌گذاری داده‌ها، و همچنین سناریوهای احتمالی که در آن‌ها برچسب‌ها در دسترس نیستند، مدل‌های خودناظارتی رویکردی ارزشمند هستند. با وجود اینکه ممکن است گاهی کمی کمتر دقیق از همتایان نظارت شده خود باشند، تجارت در کاهش تلاش‌های برچسب‌گذاری و کارایی محاسباتی آن‌ها را بسیار مفید می‌سازد.

به طور کلی، مدل‌های خودناظارتی مانند SimCLR و Barlow Twins نه تنها ارزش امتحان کردن را دارند، بلکه ابزاری ضروری در جعبه‌ابزار یادگیری ماشین مدرن هستند، به ویژه برای وظایفی که داده‌های برچسب‌گذاری شده کمیاب یا گران است. آن‌ها تعادلی مناسب بین عملکرد و عملی بودن فراهم می‌کنند و نشان می‌دهند که یادگیری خودناظارتی جایگزینی قدرتمند برای روش‌های سنتی نظارت می‌تواند باشد.

منابع و مراجع

- [1] Tabian, I., Fu, H., and Khodaei, Z. S. A convolutional neural network for impact detection and characterization of complex composite structures. *Sensors*, 19(22):4933, November 2019.
- [2] Mwiti, D. Transfer learning guide: A practical tutorial with examples for images and text in keras, April 2024. Accessed: 2024-08-28.
- [3] Pup, Federico and Atzori, Manfredo. Applications of self-supervised learning to biomedical signals: where are we now, 04 2023.
- [4] Mukherjee, S. The annotated resnet-50 - towards data science, August 2022. Accessed: 2024-08-28.
- [5] Gorlapraveen. File:vgg16.png - wikimedia commons, 2021. Accessed: 2024-08-28.
- [6] Chen, Ting, Kornblith, Simon, Norouzi, Mohammad, and Hinton, Geoffrey. A simple framework for contrastive learning of visual representations, 2020.
- [7] Tsang, S.-H. Review — simclr: A simple framework for contrastive learning of visual representations, March 2022. Accessed: 2024-08-28.
- [8] Wicaksana, P. Y. Self-supervised pre-training with simclr - prabowo yoga wicaksana - medium, March 2023. Accessed: 2024-08-28.
- [9] Zbontar, Jure, Jing, Li, Misra, Ishan, LeCun, Yann, and Deny, Stéphane. Barlow twins: Self-supervised learning via redundancy reduction, 2021.
- [10] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Deep residual learning for image recognition, 2015.

- [11] Simonyan, Karen and Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition, 2015.
- [12] Rahman, Tawsifur, Khandakar, Amith, Kadir, Muhammad Abdul, Islam, Khandaker Rejaul, Islam, Khandakar F., Mazhar, Rashid, Hamid, Tahir, Islam, Mohammad Tariqul, Kashem, Saad, Mahbub, Zaid Bin, Ayari, Mohamed Arselene, and Chowdhury, Muhammad E. H. Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization. *IEEE Access*, 8:191586–191601, 2020.
- [13] Kermany, Daniel S., Michael Goldbaum Wenjia Cai Carolina CS Valentim Huiying Liang Sally L. Baxter Alex McKeown et al. Identifying medical diagnoses and treatable diseases by image-based deep learning., 2018.
- [14] Lecun, Yann, Bengio, Yoshua, and Hinton, Geoffrey. Deep learning. *Nature*, 521(7553):436–444, May 2015. Publisher Copyright: © 2015 Macmillan Publishers Limited. All rights reserved.
- [15] Fukushima, Kunihiko. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36:193–202, 1980.
- [16] Nair, Vinod and Hinton, Geoffrey E. Rectified linear units improve restricted boltzmann machines. in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, p. 807–814, Madison, WI, USA, 2010. Omnipress.
- [17] Scherer, Dominik, Müller, Andreas, and Behnke, Sven. Evaluation of pooling operations in convolutional architectures for object recognition. in Diamantaras, Konstantinos, Duch, Wlodek, and Iliadis, Lazaros S., eds. , Artificial Neural Networks – ICANN 2010, pp. 92–101, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.
- [18] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [19] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

- [20] He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, and Sun, Jian. Deep residual learning for image recognition. in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.
- [21] Sarkar, D. A comprehensive hands-on guide to transfer learning with real-world applications in deep learning, October 2022. Accessed: 2024-08-28.
- [22] Donges, N. What is transfer learning? exploring the popular deep learning approach, August 2024. Accessed: 2024-08-28.
- [23] Zhuang, Fuzhen, Qi, Zhiyuan, Duan, Keyu, Xi, Dongbo, Zhu, Yongchun, Zhu, Hengshu, Xiong, Hui, and He, Qing. A comprehensive survey on transfer learning, 2020.
- [24] GeeksforGeeks. Self-supervised learning (ssl), December 2023. Accessed: 2024-08-28.
- [25] Balestrieri, Randall, Ibrahim, Mark, Sobal, Vlad, Morcos, Ari, Shekhar, Shashank, Goldstein, Tom, Bordes, Florian, Bardes, Adrien, Mialon, Gregoire, Tian, Yuandong, Schwarzschild, Avi, Wilson, Andrew Gordon, Geiping, Jonas, Garrido, Quentin, Fernandez, Pierre, Bar, Amir, Pirsiavash, Hamed, LeCun, Yann, and Goldblum, Micah. A cookbook of self-supervised learning, 2023.
- [26] Zhang, Richard, Isola, Phillip, and Efros, Alexei A. Colorful image colorization, 2016.
- [27] Larsson, Gustav, Maire, Michael, and Shakhnarovich, Gregory. Learning representations for automatic colorization, 2017.
- [28] He, Kaiming, Chen, Xinlei, Xie, Saining, Li, Yanghao, Dollár, Piotr, and Girshick, Ross. Masked autoencoders are scalable vision learners, 2021.
- [29] Gidaris, Spyros, Singh, Praveer, and Komodakis, Nikos. Unsupervised representation learning by predicting image rotations, 2018.
- [30] Caron, Mathilde, Bojanowski, Piotr, Joulin, Armand, and Douze, Matthijs. Deep clustering for unsupervised learning of visual features, 2019.

- [31] van den Oord, Aaron, Li, Yazhe, and Vinyals, Oriol. Representation learning with contrastive predictive coding, 2019.
- [32] Grill, Jean-Bastien, Strub, Florian, Altché, Florent, Tallec, Corentin, Richemond, Pierre H., Buchatskaya, Elena, Doersch, Carl, Pires, Bernardo Avila, Guo, Zhaohan Daniel, Azar, Mohammad Gheshlaghi, Piot, Bilal, Kavukcuoglu, Koray, Munos, Rémi, and Valko, Michal. Bootstrap your own latent: A new approach to self-supervised learning, 2020.
- [33] Chen, Xinlei and He, Kaiming. Exploring simple siamese representation learning, 2020.
- [34] Caron, Mathilde, Touvron, Hugo, Misra, Ishan, Jégou, Hervé, Mairal, Julien, Bojanowski, Piotr, and Joulin, Armand. Emerging properties in self-supervised vision transformers, 2021.
- [35] Bardes, Adrien, Ponce, Jean, and LeCun, Yann. Vicreg: Variance-invariance-covariance regularization for self-supervised learning, 2022.
- [36] Bao, Hangbo, Dong, Li, Piao, Songhao, and Wei, Furu. Beit: Bert pre-training of image transformers, 2022.