



دانشگاه‌یزد

دانشکده مهندسی کامپیووتر

پایان‌نامه جهت دریافت درجه کارشناسی ارشد

در رشته مهندسی کامپیووتر گرایش هوش مصنوعی و رباتیک

## ارزیابی خودکار کیفیت تصویر متنی با استفاده از شبکه عصبی عمیق

استاد راهنما:

دکتر علی‌محمد لطیف

استاد مشاور:

دکتر مهدی یزدیان دهکردی

نگارش:

ريحانه غفوری‌پور

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

شناسه مدرک: EP/F005

ویرایش: تابستان ۱۴۰۰

شماره صفحات: ۱ از ۱

باسمه تعالیٰ

صور تجلیسه دفاع از پایان نامه دانشجوی

مقطع کارشناسی ارشد



دانشگاه  
دانشگاه‌ها

جلسه دفاع از پایان نامه تحصیلی آقای / خانم ..... دانشجوی کارشناسی ارشد رشته / گرایش:

عنوان: ..... تحت .....

دانشجویی: ..... شماره ..... به .....»

.....»

به ارزش: ..... واحد در تاریخ / / ۱۴ با حضور اعضای هیأت داوران تشکیل:

گردید و پس از ارزیابی، اعتبار پایان نامه برای اخذ مدرک کارشناسی ارشد به صورت زیر مورد تأیید قرار گرفت:

(الف)  قبول با درجه

عالی (نمره ۱۹ تا ۲۰)

خیلی خوب (نمره ۱۸ تا ۱۸/۹۹)

خوب (نمره ۱۶ تا ۱۷/۹۹)

متوسط (نمره ۱۴ تا ۱۵/۹۹)

امضاء

مرتبه علمی

نام و نام خانوادگی

سمت

استاد/ استادان راهنمایی

استاد/ استادان مشاور

استاد داور اول

استاد داور دوم

استاد داور سوم

استاد ناظر تحصیلات

نام و نام خانوادگی استاد ناظر تحصیلات تكميلي:

امضاء و تاریخ:



## تعهد رعایت حقوق معنوی دانشگاه یزد

تحصیلات تکمیلی

اینجانب ..... دانشآموخته مقطع کارشناسی ارشد در رشته ..... گرایش  
..... که در تاریخ ..... از پایاننامه خود تحت عنوان: .....

..... با کسب درجه ..... دفاع  
نموده‌ام، شرعاً و قانوناً متعهد می‌شوم:

۱) مطالب مندرج در این پایاننامه حاصل تحقیق و پژوهش اینجانب بوده و در مواردی که از دستاوردهای علمی و پژوهشی دیگران اعم از پایاننامه، کتاب، مقاله و غیره استفاده نموده‌ام، رعایت کامل امانت را نموده و ضمن رعایت قواعد مرجع‌دهی، اقدام به ذکر آن‌ها در فهرست منابع نموده‌ام.

۲) تمام یا بخشی از این پایاننامه قبل‌اً برای دریافت هیچ مدرک تحصیلی (هم‌سطح، پایین‌تر یا بالاتر) در سایر دانشگاه‌ها و مؤسسه‌های آموزش عالی ارائه نشده است.

۳) مقالات مستخرج از این پایاننامه کاملاً حاصل کار اینجانب بوده و از هرگونه جعل داده و یا تغییر اطلاعات پرهیز نموده‌ام. همچنین با آگاهی کامل از مفad اصول اخلاق علمی، خود را ملزم به رعایت آن و عدم انجام سرقت و جعل علمی می‌دانم.

۴) از ارسال همزمان و یا تکراری مقالات مستخرج از این پایاننامه به نشریات و یا کنگره‌های گوناگون خودداری نموده و می‌نمایم.

۵) کلیه حقوق مادی و معنوی حاصل از این پایاننامه متعلق به دانشگاه یزد بوده و متعهد می‌شوم هرگونه بهره‌مندی و یا نشر دستاوردهای حاصل از این تحقیق اعم از چاپ کتاب، مقاله، ثبت اختراع و غیره (در زمان دانشجویی و یا بعد از فراغت از تحصیل) با کسب اجازه از استاد/ استادان راهنمای، مشاور و ذکر نام دانشگاه یزد باشد.

۶) در صورت اثبات تخلف (در هر زمان) مسئولیت حقوقی آن بطور کامل بر عهده اینجانب بوده و می‌پذیرم مدرک تحصیلی صادر شده توسط دانشگاه یزد از درجه اعتبار ساقط گردد و اینجانب هیچ‌گونه ادعایی نخواهم داشت.

۷) برای انجام مراحل قبل و بعد از دفاع آیین‌نامه‌ها و شیوه‌نامه‌های مربوط در وب‌سایت تحصیلات تکمیلی دانشگاه را بطور دقیق مطالعه کرده‌ام و به مفاد آن‌ها پایبند هستم.

نام و نام خانوادگی دانشجو:

امضا و تاریخ:



## وزارت علوم، تحقیقات و فناوری

### معاونت پژوهش و فناوری

به نام خدا

### منشور اخلاق پژوهش

با استعانت از خدای سبحان و با اعتقاد راسخ به اینکه عالم محضر خدادست و او همواره ناظر بر اعمال ماست و به منظور انجام شایسته پژوهش‌های اصیل، تولید دانش جدید و بهسازی زندگانی بشر، ما دانشجویان و اعضای هیأت علمی دانشگاه‌ها و پژوهشگاه‌های کشور:

- تمام تلاش خود را برای کشف حقیقت و فقط حقیقت به کار خواهیم بست و از هرگونه جعل و تحریف در فعالیت‌های علمی پرهیز می‌کنیم.
- حقوق پژوهشگران، پژوهیدگان (انسان، حیوان، گیاه و اشیاء)، سازمان‌ها و سایر صاحبان حقوق را به رسمیت می‌شناسیم و در حفظ آن می‌کوشیم.
- به مالکیت مادی و معنوی آثار پژوهشی ارج می‌نهیم، برای انجام پژوهشی اصیل اهتمام ورزیده از سرقت علمی و ارجاع نامناسب اجتناب می‌کنیم.
- ضمن پایبندی به انصا و اجتناب از هرگونه تبعیض و تعصب، در کلیه فعالیت‌های پژوهشی رهیافتی نقادانه اتخاذ می‌کنیم.
- ضمن امانتداری، از منابع و امکانات اقتصادی، انسانی و فنی موجود استفاده بهره‌ورانه می‌کنیم.
- از انتشار غیراخلاقی نتایج پژوهش نظری انتشار موازی همپوشان و چندگانه (تکه‌ای) پرهیز می‌کنیم.
- اصل محترمانه بودن و رازداری را محور تمام فعالیت‌های پژوهشی خود قرار می‌دهیم.
- در همه فعالیت‌های پژوهشی به منافع ملی توجه کرده و برای تحقق آن می‌کوشیم.
- خویش را ملزم به رعایت کلیه هنجارهای علمی رشته خود، قوانین و مقررات، سیاست‌های حرفه‌ای، سازمانی، دولتی و راهبردهای ملی در همه مراحل پژوهش می‌دانیم.
- رعایت اصول اخلاق در پژوهش را اقدامی فرهنگی می‌دانیم و به منظور بالندگی این فرهنگ، به ترویج و اشاعه آن در جامعه اهتمام می‌ورزیم.

نام و امضاء استاد راهنما:

نام و امضاء دانشجو:

## تقدیم به:

پدر و مادرم عزیزم که پیمانه عمرم لبریز از صفا، صمیمیت و مهربانی‌شان است،  
همسر همراهم که سایه مهربانی‌اش سایه‌سار زندگی‌ام است،  
نرگس نازنینم که با تولدش تکمیل این پایان‌نامه را با شیرینی و امید همراه ساخت،  
خواهرم که وجودش شادی‌بخش و مایه دلگرمی من است،  
روح پرمه ر مادربزرگم که صبر و شکرگزاری را در زندگی به من آموختند،

۹

تمامی پویندگان علم و دانش

## سپاس و قدردانی:

سپاس خداوند یکتا که رحمت و دانش او در سراسر گیتی گستردہ شده و علم و دانش حقیقی را برابر هر که خواهد موهبت فرماید. رحمت و لطف او را بینهایت سپاس می‌گوییم چه آنکه فهم و درک مطالب این پژوهش را برابر من ارزانی داشت.

از همسرم که محبت و دلگرمی‌هایش مشوق من است سپاسگزارم.

از اساتید ارجمند دکتر علی‌محمد لطیف و دکتر مهدی یزدیان که اینجانب را در تهییه این پژوهش یاری نمودند کمال قدردانی را دارم و از خداوند متعال توفیق روزافزون ایشان را خواستارم.

## چکیده

در سال‌های اخیر محتوای چندرسانه‌ای به‌ویژه تصاویر دیجیتال به لحاظ ماهیت بصری و تأثیرگذاری در زندگی روزمره انسان‌ها نقش مهمی ایفا می‌کند. موثرترین ویژگی در میزان تأثیرگذاری تصاویر، کیفیت آن است. با توجه به اهمیت کیفیت تصاویر در حوزه‌های کاربردی، ارزیابی آن مورد توجه قرار گرفته است و لزوم ارزیابی کیفیت تصویر به صورت خودکار را ایجاب می‌کند. ارزیابی کیفیت تصاویر بدون اطلاع از نوع خرابی اعمال شده بر تصویر و بدون دسترسی به تصویر مرجع یکی از چالش‌های موجود در زمینه‌ی پردازش تصاویر دیجیتال است. در این زمینه، مقایسه تصویر تحریف‌شده با تصویر مرجع امکان پذیر نیست. با توجه به توسعه و موفقیت حوزه یادگیری عمیق در ارائه روش‌های قدرتمند برای حل مسائل، این پژوهش از رویکرد یادگیری عمیق و بازنظمی شبکه عصبی کانولوشنال برای ارزیابی کیفیت تصویر به صورت خودکار استفاده گردید. مدل پیشنهادی بر اساس معماری Fastai و با استفاده از یادگیری انتقالی ارائه شده است. این پژوهش از طریق بازنظمی شبکه عصبی کانولوشنال ResNet18 که از پیش بر روی تصویر ImageNet برای مسئله کلاس‌بندی آموزش دیده شده است، برای مسئله ارزیابی کیفیت تصویر که یک مسئله رگرسیون است مدلی را ارائه می‌دهد. این مدل بر روی مجموعه داده تصویری TID2013 متشكل از ۳۰۰۰ تصویر آموزش دیده است. عملکرد مدل پیشنهادی با سایر مدل‌های پیشین ارائه شده در این مسئله مقایسه گردید. عملکرد مدل بر اساس کیفیت ارزیابی شده تصاویر با معیار ریشه میانگین مربعات خطأ برابر ۰/۵۶ است. نتایج حاصل از این پژوهش بیان‌کننده آن است که می‌توان از مدل پیشنهادی برای ارزیابی کیفیت یک تصویر به صورت خودکار در حوزه‌های کاربردی آن مانند تشخیص انحراف تصویر، بهبود تصویر و حذف عوامل تخریبی استفاده کرد.

**کلیدواژه‌ها:** ارزیابی کیفیت تصویر، مجموعه داده TID2013، یادگیری عمیق.

## فهرست مطالب

۷	<b>فصل ۱: مقدمه</b>
۸	۱-۱ بیان مسئله و ضرورت انجام آن
۹	۲-۱ هدف از تحقیق
۱۰	۳-۱ ساختار پایان نامه
۱۱	<b>فصل ۲: پیشینه تحقیق</b>
۱۴	۱-۲ ارزیابی کیفیت تصویر با مرجع کامل
۱۸	۲-۲ ارزیابی کیفیت تصویر با مرجع کاوش یافته
۲۰	۳-۲ ارزیابی کیفیت تصویر بدون مرجع
۲۵	ارزیابی کیفیت تصویر متنی
۲۸	۴-۲ جمع‌بندی
۲۹	<b>فصل ۳: روش پیشنهادی</b>
۳۰	۱-۳ تعریف دقیق مسئله
۳۱	۲-۳ روش حل مسئله
۳۱	۳-۳ مدل پیشنهادی مبتنی بر یادگیری عمیق
۳۴	۴-۳ روش آموزش مدل پیشنهادی
۳۵	۵-۳ ابر پارامترهای مدل
۳۵	۱-۵-۳ نرخ یادگیری
۳۶	۲-۵-۳ تعداد مراحل آموزش
۳۷	۳-۵-۳ تابع فعال‌ساز
۳۷	۴-۵-۳ الگوریتم بهینه‌سازی
۳۸	۵-۵-۳ تابع هزینه
۳۸	۶-۵-۳ اندازه‌ی دسته

۳۸.....	نرخ حذف تصادفی	۷-۵-۳
۳۸.....	لایه ادغام	۸-۵-۳
۳۹.....	پارامترهای مدل	۶-۳
<b>۴۱.....</b>	<b>فصل ۴: نتایج، تجزیه و تحلیل و تفسیر آنها</b>	
۴۲.....	مجموعه دادگان	۱-۴
۴۳.....	A57 مجموعه داده	۱-۱-۴
۴۴.....	LIVE مجموعه داده	۲-۱-۴
۴۴.....	CSIQ مجموعه داده	۳-۱-۴
۴۶.....	TID2013 و TID2008 مجموعه دادگان	۴-۱-۴
۵۰.....	C CID2014 و CID2013 مجموعه دادگان	۵-۱-۴
۵۲.....	KADID700K و KADID10K مجموعه دادگان	۶-۱-۴
۵۳.....	AVA مجموعه داده	۷-۱-۴
۵۵.....	NITS-IQA مجموعه داده	۸-۱-۴
۵۶.....	محیط پیاده‌سازی	۲-۴
۵۷.....	بررسی عملکرد یک معیار	۳-۴
۵۷.....	دقت پیش‌بینی	۱-۳-۴
۵۸.....	یکنواختی پیش‌بینی	۲-۳-۴
۵۹.....	نتایج حاصل از روش پیشنهادی	۴-۴
۶۰.....	بررسی عملکرد مدل آموزش‌دیده مبتنی بر روش پیشنهادی	۵-۴
۶۶.....	تحلیل عملکرد مدل پیشنهادی در نمونه کاربردی	۶-۴
<b>۷۱.....</b>	<b>فصل ۵: نتیجه‌گیری و پیشنهادها</b>	
۷۲.....	جمع‌بندی	۱-۵
۷۲.....	نتیجه‌گیری	۲-۵

۷۳	۳-۵ پیشنهادها
۷۴	منابع و مأخذ
۸۱	پیوست‌ها

## فهرست شکل‌ها

شکل ۱-۲ : نمایی از سیستم تصویربرداری زبان انسان .....	۱۲
شکل ۲-۲ : دو نمونه تصویر بی‌کیفیت و باکیفیت از زبان انسان.....	۱۲
شکل ۳-۲: نتیجه تشخیص آفاتی که بوته گوجه‌فرنگی را تحت تأثیر قرار داده است .....	۱۳
شکل ۴-۲ : دو تصویر با مقدار PSNR یکسان و وضوح متفاوت .....	۱۵
شکل ۵-۲ : مدل امیرشاھی و همکاران.....	۱۶
شکل ۶-۲ : مدل راشنانی و همکاران.....	۱۷
شکل ۷-۲ : نمایشی از تصویر مرجع، تصویر تحریفشده و تصویر غیر تراز با صحنه مشابه.....	۱۸
شکل ۸-۲ : تصاویر حاصل از لایه‌های کانولوشنال مدل لامیچه و همکاران .....	۲۲
شکل ۹-۲ : معماری مدل لی و همکاران.....	۲۴
شکل ۱۰-۲ : مدل طالبی و میلانفر .....	۲۵
شکل ۱۱-۲ : مراحل پیش‌پردازش تصویر در مدل لی کنگ و همکاران.....	۲۷
شکل ۱۲-۲ : معماری شبکه لی کنگ و همکاران .....	۲۸
شکل ۱-۳ : معماری شبکه ResNet18 .....	۳۲
شکل ۲-۳ : نمایی از مراحل پیاده‌سازی مدل پیشنهادی.....	۳۳
شکل ۳-۳ : بلوک رگرسیون مدل پیشنهادی .....	۳۳
شکل ۱-۴: ۹ نمونه تصویر اعوجاج یافته مجموعه داده A57 .....	۴۳
شکل ۲-۴ : نمونه تصاویر مجموعه داده LIVE .....	۴۴
شکل ۳-۴ : ۳۰ تصاویر مرجع مجموعه داده CSIQ .....	۴۶
شکل ۴-۴ : تصاویر مرجع مجموعه داده TID2013 .....	۴۶
شکل ۵-۴ : نمایی از سطوح مختلف چهار اعوجاج مجموعه داده TID2013 .....	۴۷
شکل ۶-۴: ۲۴ نوع اعوجاج مجموعه داده TID2013 بر روی یک تصویر .....	۴۸
شکل ۷-۴ : نمایی از نرم‌افزار نمره دهی به تصاویر TID2013 و TID2008 .....	۴۹

..... شکل ۸-۴ : نمونه‌ای از تصاویر TID2013 با مقادیر MOS متناظر	۵۰
..... شکل ۹-۴ : نمونه‌هایی از تصاویر مرجع مجموعه داده CID2013	۵۲
..... شکل ۱۰-۴ : اعوجاج انتشار رنگ در پنج سطح بر روی یک تصویر KADID10K	۵۲
..... شکل ۱۱-۴ : تصاویر مرجع مجموعه دادگان KADID10K و KADID700K	۵۳
..... شکل ۱۲-۴ : اعوجاج‌های مجموعه دادگان KADID700K و KADID10K	۵۴
..... شکل ۱۳-۴ : نمونه تصاویر مجموعه داده AVA با موضوع چالش منظره	۵۴
..... شکل ۱۴-۴ : تصاویر مرجع مجموعه داده NITS-IQA	۵۵
..... شکل ۱۵-۴ : انواع اعوجاج‌های مجموعه داده NITS-IQA	۵۶
..... شکل ۱۶-۴ : نوسان تابع هزینه برای داده‌های آموزش و اعتبارسنج	۵۹
..... شکل ۱۷-۴ : رگرسیون پیش‌بینی مدل پیشنهادی نسبت به MOS	۶۰
..... شکل ۱۸-۴ : پیش‌بینی مدل پیشنهادی بر روی ۹ تصویر آزمون مجموعه داده TID2013	۶۱
..... شکل ۱۹-۴ : عملکرد مدل پیشنهادی بر روی دو تصویر طبیعی اعوجاج یافته	۶۱
..... شکل ۲۰-۴ : عملکرد مدل پیشنهادی بر روی تصاویر KADID10K متأثر از اعوجاج تاری گوسی.	۶۳
..... شکل ۲۱-۴ : عملکرد مدل پیشنهادی بر روی تصاویر KADID10K متأثر از تغییر کنترast	۶۴
..... شکل ۲۲-۴ : عملکرد مدل پیشنهادی بر روی دو تصویر متنی متأثر از تاری فوکوس	۶۵
..... شکل ۲۳-۴ : عملکرد مدل پیشنهادی بر روی تصاویر متنی متأثر از اعوجاج تاری و نویز گوسی.	۶۵
..... شکل ۲۴-۴ : عملکرد مدل پیشنهادی بر روی تصاویر پرسنلی	۶۸
..... شکل ۲۵-۴ : عملکرد مدل پیشنهادی بر روی تصاویر کارت ملی	۶۸
..... شکل ۲۶-۴ : عملکرد مدل پیشنهادی بر روی تصاویر شناسنامه	۶۹
..... شکل ۱-۵ : مدل پیشنهادی	۷۳

## فهرست جداول

جدول ۱-۲ : ارزیابی مدل لامیچه و همکاران .....	۲۲
جدول ۲-۲ : ارزیابی مدل طالبی و میلانفر بر اساس معیار SROCC	۲۵
جدول ۱-۳ : مقادیر ابر پارامترهای مدل پیشنهادی	۳۹
جدول ۲-۳ : تعداد پارامترهای مدل پیشنهادی	۴۰
جدول ۱-۴ : اعوجاج‌های مجموعه دادگان A57, LIVE, NITS-IQA و CSIQ	۴۵
جدول ۲-۴ : لیست اعوجاج‌های اعمال شده بر روی مجموعه دادگان TID2008 و TID2013	۵۱
جدول ۳-۴ : مجموعه دادگان حوزه ارزیابی کیفیت تصویر	۵۶
جدول ۴-۴ : نتایج ارزیابی مدل پیشنهادی	۶۰
جدول ۵-۴ : عملکرد سه نسخه ResNet در ارزیابی کیفیت تصویر	۶۶
جدول ۶-۴ : مقایسه مدل پیشنهادی با سایر روش‌ها بر روی پایگاه داده TID2013	۶۶

# فصل ۱:

مقدمه

با رشد روزافزون دوربین‌های دیجیتال و تلفن‌های همراه هوشمند، محتوای چندرسانه‌ای به‌ویژه تصاویر دیجیتال به یک رسانه مهم برای انتقال اطلاعات و برقراری ارتباط تبدیل شده است. ثبت و نگهداری تصاویر به لحاظ ماهیت بصری و تأثیرگذاری آن در زمینه‌های مختلف مهم است. دستگاه‌های تصویربرداری، سیستم‌های ذخیره‌سازی، الگوریتم‌های کدگذاری و انتقال اطلاعات می‌توانند باعث ایجاد اعوجاج‌هایی شوند که بر کیفیت تصویر تأثیر می‌گذارند [۱]. تصاویر دیجیتال در فرایند اخذ ممکن است به دلیل نقص سنسور دوربین یا انتقال از طریق یک کانال نویزدار و یا با اعمال برخی فرآیندها، تحریف شوند. توانایی اندازه‌گیری قابل اعتماد کیفیت تصویر یک عنصر کلیدی است.

## ۱- بیان مسئله و ضرورت انجام آن

در سال‌های اخیر با توجه به اهمیت کیفیت تصاویر در حوزه‌های کاربردی، ارزیابی آن مورد توجه قرار گرفته است. تخریب تصویر باعث کاهش توانایی انسان در استخراج و درک اطلاعات می‌شود. شناسایی و تعیین کمیت اعوجاج تصویر بهمنظور تائید، کنترل و بهبود کیفیت تصویر اهمیت دارد. از طرف دیگر، ارزیابی کیفیت تصویر نقش مهمی در ارزیابی هر سختافزار و نرمافزار جدید، تکنیک‌های اکتساب تصویر، الگوریتم‌های بازسازی و پس پردازش تصویر ایفا می‌کند. پیش‌بینی کیفیت تصاویر دیجیتال به مدیریت تصاویر باکیفیت پایین در طول فرآیند ضبط و یا افزایش کیفیت تصویر با تکیه بر برخی روش‌های بازیابی تصویر کمک می‌کند.

رویکردهای ارزیابی کیفیت تصویر نقش اصلی را در ارتباط و پردازش سیگنال بصری ایفا می‌کنند. انسان‌ها به عنوان کاربران نهایی تصاویر بهترین قضاوت کننده کیفیت بصری هستند. کیفیت بصری یک تصویر که توسط انسان ارزیابی می‌شود، همیشه به عنوان دقیق‌ترین و مطمئن‌ترین روش در نظر گرفته می‌شود. به دلیل ماهیت وقت‌گیر و پرهزینه بودن نیروی انسانی، برای ارزیابی بلاذرنگ کیفیت تصویر از روش ناظر انسانی کمتر استفاده می‌شود. حتی در شرایطی که هزینه و زمان نگران‌کننده نیست، نتایج اندازه‌گیری به تجربه قبلی و شرایط روانی فرد ارزیاب بستگی دارد. سیستم ارزیابی کیفیت تصویر مطلوب مدلی است که قادر به پیش‌بینی کیفیت تصویر به طور

دقیق و مطابق با ادراک انسان است. بسیاری از محققان سعی کرده‌اند یک معیار کیفیت همه‌منظوره را تعریف کنند که تا حد امکان نزدیک به قضاوت انسان باشد و کیفیت ادراکی یک تصویر را ارائه دهد [۹-۲].

در سال‌های اخیر، شبکه‌های عصبی مختلفی برای ارزیابی کیفیت تصویر ارائه شده است [۱۰-۱۴]. به دلیل اهمیت کاربردی تصویر، بهبود کارایی مدل‌های خودکار ارزیاب کیفیت تصویر حائز اهمیت است و می‌تواند در کارکرد حوزه‌های کاربردی تصویر از جمله ارزیابی الگوریتم‌های حذف نویز مؤثر واقع شود.

## ۲-۱ هدف از تحقیق

هدف از این پژوهش ارائه یک مدل ارزیابی خودکار کیفیت تصویر مبتنی بر یادگیری عمیق و بازنظمی شبکه عصبی عمیق است. این تحقیق در صدد آن است با انتخاب بستر پیاده‌سازی مناسب که نتیجه مستقیم بر روی عملکرد مدل دارد، یک شبکه عصبی عمیق را که از پیش آموزش دیده برای مسئله رگرسیون ارزیابی کیفیت تصویر بدون تصویر مرجع بازنظمی کند و عملکرد آن را با سایر روش‌ها و مدل‌های پیشین مقایسه کند.

رونده کار این تحقیق بدین شرح است که در ابتدا بستر Fastai و زبان برنامه‌نویسی پایتون برای پیاده‌سازی مدل انتخاب شد. جهت به کارگیری رویکرد یادگیری عمیق، از شبکه ResNet18 که بر روی پایگاه داده ImageNet برای مسئله کلاس‌بندی آموزش دیده است، استفاده گردید. شبکه ResNet18 بازنظمی شد و یک بلوك رگرسیون برای مسئله پژوهش طراحی و پیاده‌سازی شد و به انتهای ساختار شبکه پایه اضافه شد. مدل پیشنهادی بر روی مجموعه داده TID2013 آموزش داده شد. مجموعه داده TID2013 شامل ۳۰۰۰ تصویر است. هر تصویر توسط میانگین نظرات ۵۰۰ انسان در مورد کیفیت آن، برچسب‌گذاری شده است. این شبکه بر روی تصاویر جدید مورد آزمایش قرار گرفت و عملکرد مدل پیشنهادی منجر به مقدار ریشه میانگین مربعات خطای  $0.54$  شد.

<sup>1</sup>Backbone

### ۳-۱ ساختار پایان نامه

پایان نامه، به این شرح سازمان دهی شده است: در فصل دوم به مروری بر منابع و مطالعات انجام شده در حوزه تحقیق پرداخته می شود. فصل سوم روش پیشنهادی مبتنی بر رویکرد یادگیری عمیق را مطرح می کند. فصل چهارم به بررسی مجموعه دادگان موجود در حوزه ارزیابی کیفیت تصویر و عملکرد روش پیشنهادی مطرح شده در فصل چهارم می پردازد. در فصل پنجم نتیجه گیری صورت می گیرد و پیشنهادهایی برای تحقیق های آینده بیان می شود.

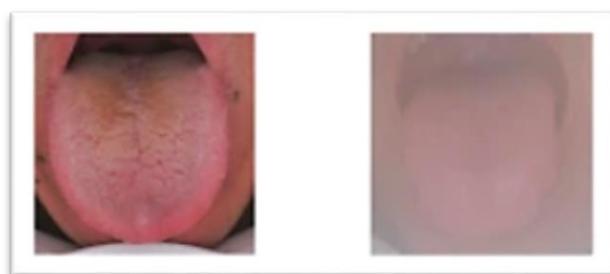
## **فصل ۲:**

**پیشینه تحقیق**

در این فصل مطالعات انجام شده در حوزه ارزیابی کیفیت تصویر بررسی می‌شوند. تشخیص بیماری‌های خاص از طریق تصاویر زبان انسان، یک نمونه کاربردی تصاویر در پزشکی است. سیستمی که برای عکس‌برداری از زبان طراحی می‌شود ممکن است دچار اعوجاج شود و تصویر گرفته شده از کیفیت مطلوب بی‌برخوردار نباشد. از آنجا که دقت و کیفیت تصاویر پزشکی در تشخیص و هزینه‌های درمان مؤثر است، لازم است ارزیابی کیفیت تصاویر به صورت خودکار صورت گیرد. شکل ۱-۲ نمایی از سیستم تصویربرداری از زبان را نشان می‌دهد. یک نمونه تصویر با کیفیت و یک نمونه تصویر بی‌کیفیت که توسط سیستم تصویربرداری از زبان انسان اخذ شده است در شکل ۲-۲ آمده است.



شکل ۱-۲ : نمایی از سیستم تصویربرداری زبان انسان [۱۵]

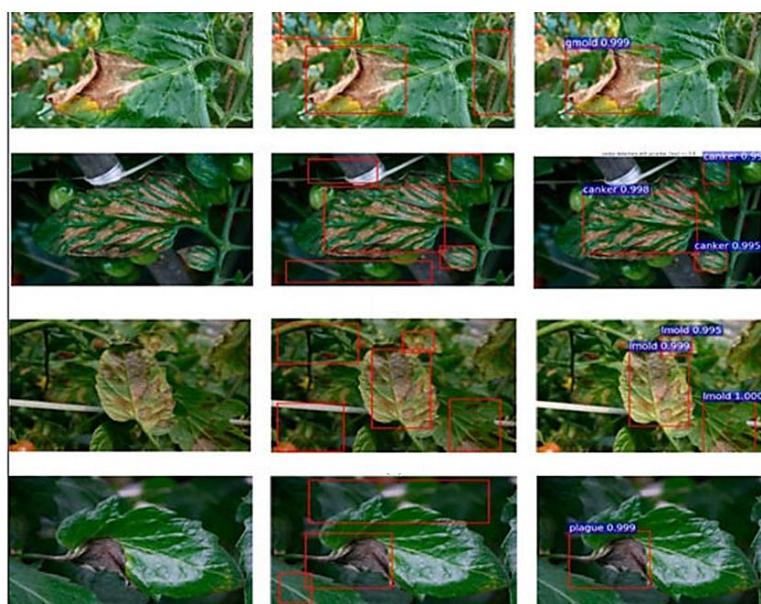


شکل ۲-۲ : دو نمونه تصویر بی‌کیفیت و با کیفیت از زبان انسان [۱۵]

تصاویر توموگرافی قفسه سینه که به طور معمول در تشخیص بالینی استفاده می‌شوند، مثال دیگری از لزوم اهمیت تصاویر در پزشکی است. ارزیابی کیفیت تصاویر توموگرافی قفسه سینه نقش مهمی در معاینه فرد بیمار ایفا می‌کند. ارزیابی کیفیت این تصاویر هنوز یک فرآیند دستی و ذهنی

است و حتی رادیولوژیست‌های با تجربه به دلیل خستگی، سوگیری‌های ادراکی و سوگیری‌های شناختی اشتباه می‌کنند. خودکارسازی فرایند ارزیابی کیفیت تصاویر توموگرافی می‌تواند یک ارزیابی عینی قابل اعتماد نتیجه دهد [۱۶].

یک نمونه دیگر از کاربرد تصاویر در حوزه کشاورزی است. بیماری‌های زراعی یک تهدید بزرگ برای امنیت غذایی هستند و شناسایی سریع مهم است. بیماری‌های گیاهی آثار خاصی بر روی ساختار برگی گیاه بر جای می‌گذارند. شناسایی سریع این بیماری‌ها به دلیل نبود زیرساخت‌های لازم دشوار است. پیشرفت‌های اخیر در بینایی کامپیوتر و افزایش نفوذ تلفن‌های هوشمند راه را برای شناسایی بیماری‌های زراعی هموار کرده است و می‌توان از روی تصویر برگ درختان آسیب‌شناسی نمود. شکل ۳-۲ چهار نمونه از شناسایی آفت بوته گوجه‌فرنگی از روی تصاویری از برگ آن نشان می‌دهد. چالشی که در اینجا مطرح است این است که تصاویر گرفته شده از برگ درختان ممکن است حاوی نویز و تاری باشد که منجر به تشخیص نادرست شود. پیاده‌سازی برنامه حذف نویز و تاری بر بستر سخت‌افزاری تلفن همراه ممکن نیست. رویکرد بهینه این است که پیش از اقدام به آسیب‌شناسی زراعی، تصویر دارای اعوجاج شناسایی شود. ارزیابی کیفیت تصاویر گرفته شده جهت انتخاب تصویر قابل استناد برای آسیب‌شناسی درست لازم است.



شکل ۳-۲ : نتیجه تشخیص آفاتی که بوته گوجه‌فرنگی را تحت تأثیر قرار داده است [۱۷]

در سال‌های اخیر، روش‌های مختلفی برای ارزیابی تصاویر طبیعی ارائه شده است. ارزیابی کیفیت تصویر توسط ناظر انسانی می‌تواند صورت پذیرد که به این نوع ارزیابی، سنجش ادراکی کیفیت تصویر<sup>۱</sup> گفته می‌شود. این روش ارزیابی زمان‌بر و پرهزینه است. به جهت صرفه‌جویی در هزینه و تسريع فرایند ارزیابی، روش‌های سنجش عینی کیفیت تصاویر<sup>۲</sup> به وجود آمدند. این روش‌ها را می‌توان با توجه به در دسترس بودن تصویر مرجع و حجم اطلاعات موجود در مرحله ارزیابی به سه دسته، روش ارزیابی کیفیت تصویر با مرجع کامل<sup>۳</sup>، روش ارزیابی کیفیت با تصویر مرجع کاهش‌یافته<sup>۴</sup> و روش ارزیابی کیفیت بدون تصویر مرجع<sup>۵</sup> طبقه‌بندی کرد [۱۸].

## ۱-۲ ارزیابی کیفیت تصویر با مرجع کامل

در روش FR-IQA، تصویر اصلی متناظر تصویر اعوجاج یافته به عنوان تصویر مرجع قابل دسترس است و می‌توان آن را با تصویر اعوجاج یافته برای محاسبه تفاوت آن‌ها مقایسه کرد. از جمله الگوریتم‌های رایج این حوزه می‌توان به میانگین مربعات خطأ<sup>۶</sup> میانگین خطای قدر مطلق<sup>۷</sup> و حداقل نسبت سیگنال به نویز<sup>۸</sup> اشاره کرد که به صورت رابطه (۱-۳)، (۲-۳) و (۳-۳) بیان می‌شوند [۱۸].

$$MSE = \sum_{I=1}^n \frac{(f(i, j) - g(i, j))^2}{n} \quad (1-3)$$

$$MAD = \sum_{I=1}^n \frac{|f(i, j) - g(i, j)|}{n} \quad (2-3)$$

$$PSNR = 10 \times \log \frac{255}{MSE} \quad (3-3)$$

n در رابطه (۱-۳) و (۲-۳) تعداد تصاویر اعوجاج یافته،  $f(i, j)$  تصویر اصلی و  $g(i, j)$  تصویر اعوجاج یافته است. هر چه تصویر اعوجاج یافته به تصویر اصلی بیش‌تر شبیه باشد، مقدار MAD و MSE کم‌تر

<sup>۱</sup>Subjective Image Quality Assessment

<sup>۲</sup>Objective Image Quality Assessment

<sup>۳</sup>Full Reference Image Quality Assessment (FR-IQA)

<sup>۴</sup>Reduced Reference Image Quality Assessment (RR-IQA)

<sup>۵</sup>No Reference Image Quality Assessment (NR-IQA)

<sup>۶</sup>Mean Squared Error (MSE)

<sup>۷</sup>Mean Absolute Difference (MAD)

<sup>۸</sup>Peak Signal to Noise Ratio (PSNR)

و مقدار PSNR بیشتر خواهد بود. اگرچه این سه معیار به طور گسترده برای اندازه‌گیری کیفیت تصاویر مورد استفاده قرار می‌گیرند، نتایج آن‌ها با سیستم بینایی انسان سازگار نیست. در شکل ۴-۲ دو تصویر را نشان می‌دهد که با وجود داشتن وضوح متفاوت، مقدار PSNR آن‌ها یکسان و برابر با ۲۵/۱۱db است [۱۹].



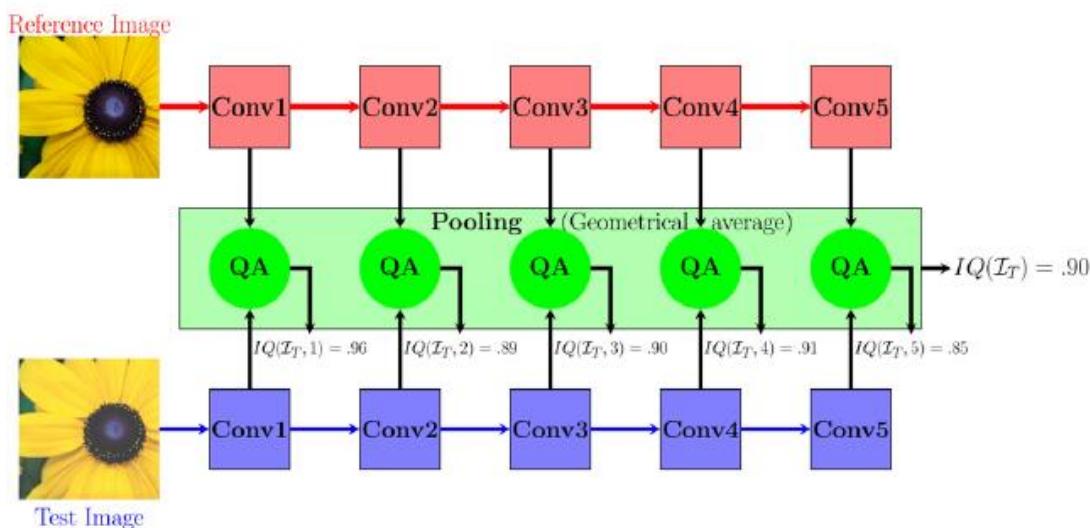
شکل ۴-۲ : دو تصویر با مقدار PSNR یکسان و وضوح متفاوت [۱۹]

وانگ و همکاران [۸] با این فرض که ادراک بصری انسان برای استخراج اطلاعات ساختاری از یک صحنه سازگار است، یک چارچوب مکمل جایگزین برای ارزیابی کیفیت تصویر بر اساس تخریب اطلاعات ساختاری معرفی کردند. در واقع آن‌ها یک شاخص تشابه ساختاری<sup>۱</sup> را پیشنهاد کردند که اطلاعات ساختاری یک تصویر را متمرکز و به کار می‌برد. ژانگ و همکاران [۹] با استفاده از این فرض که انسان‌ها تصاویر را با ویژگی‌های سطح پایین، به عنوان مثال، خطوط درک می‌کنند، یک شاخص شباهت ویژگی پیشنهاد کرد. چنین الگوریتم‌هایی روی تصاویر واقعی ضعیف عمل می‌کنند و با ارزیابی ذهنی بینایی انسان که ویژگی‌های انتزاعی تصویر استخراج می‌شوند؛ مطابقت ندارند [۲۰, ۲۱].

با ظهور رویکردهای یادگیری ماشین، معیارهای جدید کیفیت تصویر بر اساس یادگیری ماشین با عملکرد بهتر نسبت به معیارهای معروفی چون حداقل نسبت سیگنال به نویز و SSIM

<sup>۱</sup>Structural Similarity Index Measure (SSIM)

توسعه یافتند. در پژوهش امیرشاھی و همکاران [۱۰]، یک روش جدید برای ارزیابی کیفیت تصویر مبنی بر رویکرد FR-IQA است که با استفاده از ویژگی‌های تصویر استخراج شده از مدل شبکه عصبی کانولوشنال AlexNet، ارائه شده است. در مدل امیرشاھی و همکاران نقشه‌های ویژگی در چندین لایه استخراج می‌گردد و شباهت ویژگی‌های آن‌ها در هر لایه مقایسه می‌شود. سپس امتیازهای شباهت در بین لایه‌ها جمع می‌شوند تا یک مقدار کیفیت کلی به دست آید. نتایج نشان می‌دهد که ویژگی‌های CNN در سطوح چندگانه نسبت به ویژگی‌های دستی در سایر روش‌های ارزیابی کیفیت تصویر برتری دارند. شکل ۵-۲ نمایی از مدل امیرشاھی و همکاران را نشان می‌دهد. در شکل ۲ IQ ( $I_{T; n}$ ) مقدار کیفیت تصویر در لایه کانولوشنال  $n$  ام و IQ ( $I_T$ ) مقدار نهایی کیفیت تصویر ورودی را نشان می‌دهد. IQ ( $I_T$ ) از طریق میانگین هندسی مقادیر کیفیت تصویر با توجه به تصویر مرجع در هر لایه کانولوشن حاصل می‌شود. نقشه‌های ویژگی در هر پنج لایه کانولوشن استخراج و با یک معیار کیفیت مبتنی بر هیستوگرام مقایسه می‌شوند.

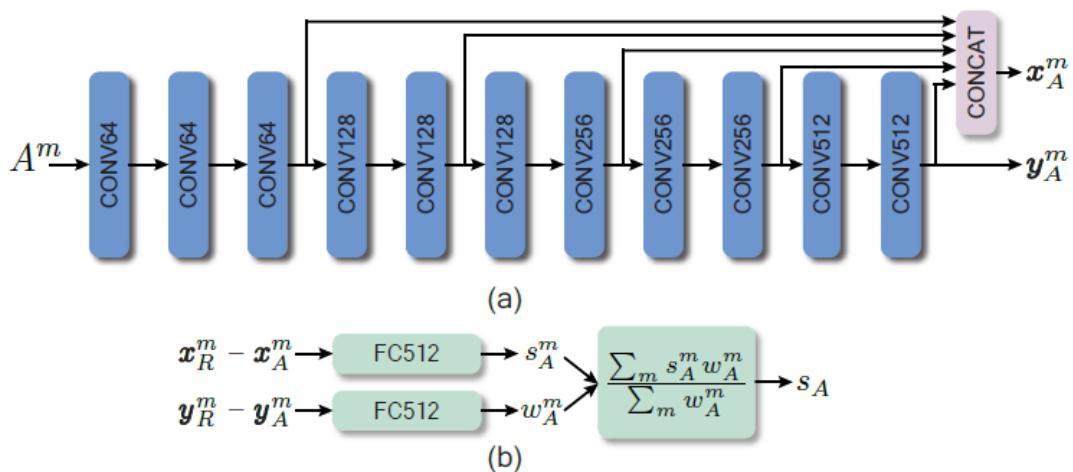


شکل ۵-۲ : مدل امیرشاھی و همکاران [۱۰]

راشنانی و همکاران [۱۳] یک مدل مبنی بر شبکه عصبی کانولوشنی مبنی بر FR-IQA پیشنهاد کردند. مدل آن‌ها با استفاده از یازده لایه کانولوشنال به همراه اتصال‌های میانبر و دو لایه

کاملاً متصل برای پیش‌بینی میزان کیفیت طراحی شده است. شکل ۶-۲ معماری مدل راشنانی و همکاران را نشان می‌دهد.

در مدل راشنانی و همکاران ابتدا تصویر ورودی به قسمت‌هایی برش داده می‌شود و به عنوان ورودی بلوک استخراج ویژگی در نظر گرفته می‌شود. بلوک استخراج ویژگی شامل ۱۱ لایه کانولوشن به همراه اتصال‌های میانبر است. هر لایه کانولوشن دارای فیلترهای  $3 \times 3$  و یک تابع فعال‌ساز ReLU غیرخطی است که پس از هر لایه فرد، ادغام حداقل  $2 \times 2$  قرار دارد (شکل ۶-۲ قسمت a). واحد محاسبه امتیاز کیفیت از دو شبکه کاملاً متصل که هر کدام دارای ۱ لایه پنهان با ۵۱۲ نورون است، به منظور محاسبه وزن‌ها و خطاهای قسمت‌های برش خورده تصویر تشکیل شده است. درنهایت میانگین وزنی امتیاز‌های همه قسمت‌های تصویر، کیفیت نهایی تصویر را نتیجه می‌دهد (شکل ۶-۲ قسمت b). مدل راشنانی و همکاران بر روی مجموعه داده TID2013 ارزیابی شد. بر اساس دو معیار ارزیابی PLCC و SROCC به ترتیب مقادیر  $0.945$  و  $0.946$  حاصل گردید.



شکل ۶-۲: مدل راشنانی و همکاران [۱۳]

در شکل ۶-۲  $A^m$  نشانگر قسمت  $m$  ام تصویر است که به عنوان ورودی وارد شبکه شده است و طی گذر از لایه‌های کانولوشن بردار ویژگی نهایی  $y_A^m$  و بردار ویژگی حاصل از ترکیب خروجی‌های ۵ لایه کانولوشنال  $x_A^m$  حاصل می‌شود.

## ۲-۲ ارزیابی کیفیت تصویر با مرجع کاهش یافته

در روش RR-IQA، تصویر اصلی متناظر تصویر تحریف شده در دسترس نیست و فقط بخشی از اطلاعات در مورد تصویر مرجع، با استفاده از یک کanal کمکی و یا با گنجاندن برخی اطلاعات در تصویر تحریف شده موجود است. این روش را می‌توان به سه گروه طبقه‌بندی کرد [۲۲].

گروه اول از ویژگی‌های مبتنی بر آمار صحنه طبیعی<sup>۱</sup> استفاده می‌کند که توزیع خروجی فیلترهای تصویر را رمزگذاری می‌کند. روش‌های مبتنی بر آمار صحنه طبیعی می‌توانند ویژگی‌های تصویر را از حوزه تبدیل مثل تبدیل موجک و تبدیل کسینوس گستته و یا از حوزه مکان استخراج کنند [۲۳-۲۶].

لیانگ و همکاران [۲۷] یک شبکه عصبی کانولوشنال عمیق دومسیره<sup>۲</sup> را پیشنهاد کردند که از یک تصویر غیر تراز با صحنه مشابه با تصویر اصلی برای ارزیابی کیفیت تصویر نیاز آن‌ها ارائه کردند در راستای حل این اشکال بود که بسیاری از روش‌های ارزیابی کیفیت تصویر نیاز دارند که تصویر مرجع از نظر پیکسلی با تصویر تحریف شده هم‌راستا باشد. در نتیجه با مدل پیشنهادی لیانگ و همکاران می‌توان از تصویر غیر تراز با صحنه مشابه به خوبی برای مرجع استفاده کرد.



شکل ۷-۲: نمایشی از تصویر مرجع، تصویر تحریف شده و تصویر غیر تراز با صحنه مشابه [۲۷]

<sup>۱</sup>Natural Scene Statistics (NSS)

<sup>۲</sup>Dual Path Deep Convolutional Neural Network (DCNN)

در شکل ۷-۲ از بالا به پایین، ردیف اول تصاویر مرجع را نشان می‌دهد که در دسترس نیست.

ردیف دوم تصاویر تحریف شده و ردیف سوم تصاویر مشابه با تصاویر مرجع که با هیچ تغییر هندسی به تصاویر تحریف شده مرتبط نمی‌شوند را نمایش می‌دهد.

گروه دوم نیاز به دانش در مورد فرآیند اعوجاج مانند فشرده‌سازی تصویر دارد. چونو و همکاران

[۲۸] یک طرح ارزیابی کیفیت تصویر با توجه به مرجع کاهش‌یافته و با استفاده از کدگذاری منبع توزیع شده ارائه کردند. در مدل آن‌ها ابتدا بردار ویژگی از تصویر اصلی استخراج می‌شود و تنها جریان بیت Slepian-Wolf خود را منتقل می‌کند. با انتخاب مناسب نرخ برای جریان بیت با توجه به کیفیت تصویر مجاز، یک گیرنده می‌تواند به درستی بردار ویژگی را با استفاده از تصویر دریافتی، به عنوان اطلاعات جانبی، بازسازی کند. نتایج شبیه‌سازی نشان می‌دهد که بردار ویژگی را می‌توان به طور مؤثر با کدگذاری منبع توزیع شده تا ۲۵ درصد بهتر از کدگذاری منبع معمولی فشرده کرد. در نهایت گیرنده می‌تواند کیفیت تصویر دریافتی را با استفاده از بردار ویژگی بازسازی شده تعیین کند.

گروه سوم از مدل‌های گیرنده تصویر استفاده می‌کند که فرآیند فیزیکی بینایی انسان را در بر می‌گیرد تا ارزیابی کیفیت تصویر را تسهیل کند.

در پژوهش [۲۹] یک مدل مولد داخلی که درک صحنه توسط مغز را تقلید می‌کند، بر روی نمایش پراکنده تصاویر مرجع و تحریف شده ساخته شده است. تفاوت بین آنتروپی اختلاف میزان کیفیت پیش‌بینی شده برای تصاویر مرجع و تحریف شده به عنوان شاخص کیفیت تعریف می‌شود. بر این اساس یک معیار کیفیت تصویر پیشنهاد شده است که اصل انرژی آزاد<sup>۱</sup> و شاخص مبتنی بر نمایش پراکنده برای ارزیابی کیفیت تصویر نامیده می‌شود. FSI متعلق به روش‌های RR-IQA است و برای تخمین کیفیت تصویر فقط به یک عدد از تصویر مرجع نیاز دارد.

در پژوهش [۳۰]، یک رویکرد مبنی بر اختلاف آنتروپی‌ها با مرجع کاهش‌یافته<sup>۲</sup> پیشنهاد شده است که اندازه‌گیری کیفیت از طریق اختلاف بین آنتروپی ضرایب موجک<sup>۳</sup> تصاویر مرجع و تصاویر

<sup>۱</sup>Free-Energy Principle

<sup>۲</sup>Free-Energy Principle and Sparse Representation-Based Index for Image Quality Assessment (FSI)

<sup>۳</sup>Reduced Reference Entropies Differencing (RRED)

<sup>۴</sup>Wavelet Coefficients

تحریف شده صورت می‌گیرد. به دنبال دستاورد قابل توجه SSIM، اطلاعات ساختاری به عنوان یکی از عوامل کلیدی شناخته شده است و تاکنون بسیاری از معیارهای کیفیت تصویر را برانگیخته است. گو و همکاران [۳۱] یک مدل تخرب ساختاری با در نظر گرفتن اندازه‌گیری شاخص تشابه ساختاری پیشنهاد کردند که در آن امتیاز کیفیت یک تصویر به عنوان یک ماشین بردار پشتیبان مبنی بر ادغام فاصله بین اطلاعات تخرب ساختاری تصاویر مرجع و تصاویر تحریف شده به دست می‌آید.

### ۲-۲ ارزیابی کیفیت تصویر بدون مرجع

در روش NR-IQA، تصویر اصلی در دسترس نیست و ارزیابی کیفیت تصویر اعوجاج یافته باید بدون توجه به تصویر اصلی صورت گیرد. ارزیابی کیفیت تصویر بدون مرجع چالش‌برانگیزتر از روش‌های FR-IQA است. این روش ارزیابی مانند سیستم بینایی انسان که به هیچ مرجعی برای تعیین کیفیت تصویر نیاز ندارد عمل می‌کند. هدف ارزیابی کیفیت تصویر در این روش، به حداقل رساندن همبستگی بین پیش‌بینی کیفیت تصویر و میانگین نمرات انسانی برای کیفیت تصویر است. روش‌های موجود در این دسته را می‌توان به دو زیر دسته آگاه از خرابی خاص یا مستقل و ناآگاه از خرابی تقسیم نمود. روش‌های آگاه از خرابی با آگاهی از نوع اعوجاج، کیفیت یک تصویر را پیش‌بینی می‌کنند. این الگوریتم‌ها می‌توانند یک یا چند اعوجاج در یک تصویر را اندازه‌گیری کنند. به دلیل اینکه اطلاعات در مورد نوع اعوجاج در بیش‌تر برنامه‌های کاربردی در دسترس نیست و یا دامنه اعوجاج‌ها گسترده است، کاربرد این الگوریتم‌ها محدود است.

الگوریتم‌های ناآگاه از خرابی بر روی اعوجاج خاصی متتمرکز نمی‌شوند و می‌توانند بر روی انواع مختلف اعوجاج‌ها اعمال شوند. این نوع الگوریتم‌ها وظیفه پیچیده‌تری نسبت به دسته قبل دارند؛ زیرا اطلاعاتی در مورد نوع اعوجاج در دسترس نیست و تنها با توجه به تصویر اعوجاج یافته و اطلاعات موجود در آن کیفیت تصویر را ارزیابی می‌کنند. تکنیک‌های مختلفی در حوزه الگوریتم‌های ناآگاه از خرابی ارائه شده است که در ادامه به آن پرداخته می‌شود.

لی در مرجع [۳۲]، پیشنهاد می‌کند که کیفیت تصویر با سه معیار سطح وضوح لبه، سطح نویز تصادفی و سطح نویز ساختاری ارزیابی شود. او معتقد است این سه معیار به طور مشترک یک

رویکرد اکتشافی برای توصیف مهم‌ترین جنبه‌های کیفیت بصری هستند و ابزارهای مختلف ریاضی را برای تخمین دقیق و قوی این سه معیار بررسی می‌کند.

رویکردهای یادگیری ماشین مبتنی بر ارزیابی کیفیت بدون مرجع که اغلب از نمرات ناظر انسانی در مرحله یادگیری شبکه عصبی استفاده می‌کنند، ویژگی‌های سطح پایین [۳۳-۳۵] را از تصاویر تخریب شده استخراج می‌کنند. سپس از رگرسیون برای پیش‌بینی میزان کیفیت استفاده می‌شود. مورثی و بوویک [۳۵] روشی را برای ارزیابی کیفیت تصویر با شناسایی نوع اعوجاج آن پیشنهاد کردند. آن‌ها با توجه به اینکه ویژگی‌های آماری تصویر زمانی که اعوجاج در تصویر وجود داشته باشد تغییر می‌کند؛ با استخراج ویژگی‌هایی که با ادراک انسان مرتبط است، کیفیت تصویر را ارزیابی می‌کند. این رویکرد بر روی مجموعه داده LIVE [۳۶] ارزیابی شد و مقدار RMSE آن، ۰/۹۰ به دست آمد.

خراسادی‌زاده و همکاران [۱۸] در تلاشی یک مدل ارزیابی کیفیت تصویر بر اساس آنتروپی محلی و انحراف معیار در وضوح چند مقیاسی خود تصویر و تصویر گرادیان آن پیشنهاد کردند. این ویژگی‌های محلی در بلوک‌های  $8 \times 8$  غیر همپوشانی استخراج می‌شوند؛ سپس توزیع این مقادیر توسط یک مدل مخلوط گوسی توصیف می‌شود. به منظور اندازه‌گیری کیفیت تصویر و شناسایی نوع اعوجاج یک شبکه عصبی که با پارامترهای مدل مخلوط گوسی مقداردهی اولیه شده است آموزش داده می‌شود. رویکرد خراسادی‌زاده و همکاران بر روی مجموعه داده LIVE پیاده‌سازی و ارزیابی شد و معیار ارزیابی RMSE ۶/۹۲ حاصل گردید.

لامیچه و همکاران [۳۷] یک معیار کیفیت جدید برای تصاویر دو بعدی ارائه کردند. مدل لامیچه و همکاران از ویژگی‌های سیستم بصری انسان و ویژگی آماری صحنه‌های طبیعی استفاده می‌کند. از اطلاعات بر جسته برای شناسایی مناطق ادراکی مرتبط با کیفیت تصویر استفاده می‌شود و ویژگی‌های آمار صحنه طبیعی امکان ارزیابی تغییرهای تصویر از حالت طبیعی را با وجود اعوجاج‌های در آن فراهم می‌کند. روش لامیچه و همکاران بر روی سه مجموعه داده TID2013، CSIQ و LIVE آموزش داده شد. نتیجه ارزیابی مدل آن‌ها در جدول ۱-۲ بر اساس دو معیار ارزیابی SROCC و

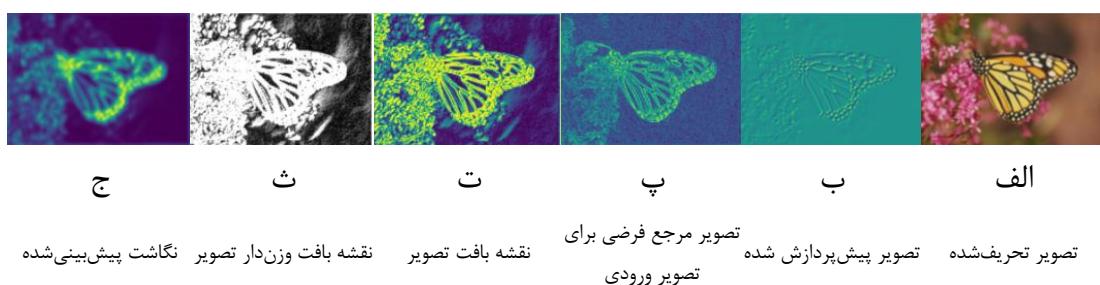
PLCC آمده است.

## جدول ۱-۲: ارزیابی مدل لامیچه و همکاران

CSIQ	TID2013	LIVE	
•/৯৬৭	•/৯১০	•/৯৭৭	SROCC
•/৯৭২	•/৯১৮	•/৯৮৮	PLCC

شکل ۲-۲ تصاویر حاصل از لایه‌های کانولوشنال در مدل ارائه شده لامیچه و همکاران را نشان می‌دهد. بر روی تصویر اعوجاج یافته پیش‌پردازش صورت می‌گیرد؛ به این صورت که یک نسخه فیلتر شده پایین گذر از طریق سه مرحله که عبارت است از اعمال فیلتر گوسی، نمونه‌برداری فرعی با ضریب ۴ و درون‌پایه، به ابعاد اصلی، به دست می‌آید.

به کمک رویکردهای یادگیری ماشین می‌توان فیلترهای تصویر دیجیتال را برای به دست آوردن ویژگی‌های خاص تصویر اعوجاج یافته به عنوان مثال لبه‌ها طراحی کرد و از آن‌ها برای ارزیابی کیفیت تصویر استفاده کرد. چنین رویکردهایی به ارزیابی مناسبی در برخی از انواع خاص تصاویر دست یافته‌اند؛ اما اشکال کار این است که آن‌ها از ویژگی‌های تصویری از پیش طراحی شده به عنوان مثال، دست‌ساز استفاده می‌کنند که ممکن است شباهت زیادی به آنچه در بینایی انسان استفاده می‌شود، نباشد. در سال‌های اخیر شبکه‌های عصبی کانولوشنال به دلیل توانایی یادگیری تصویر توجه بیشتری را به خود جلب کرده‌اند.



شکل ۸-۲: تصاویر حاصل از لایه‌های کانولوشنال مدل لامیچه و همکاران [۳۷]

یان و همکاران [۱۴] یک شبکه عصبی عمیق جدید برای پیش‌بینی کیفیت تصویر به طور دقیق بدون تکیه بر تصویر مرجع ارائه کردند. آن‌ها برای یادگیری بازنمایش ویژگی‌های مؤثرتر با رویکرد NR-IQA، یک شبکه کانولوشن دو جریانی را پیشنهاد کردند که شامل دو جزء فرعی برای تصویر و تصویر گرادیان است. انگیزه این طراحی استفاده از یک طرح دو جریانی برای گرفتن اطلاعات سطوح مختلف ورودی‌ها و کاهش دشواری استخراج ویژگی‌ها از یک جریان است. جریان گرادیان بر استخراج جزئیات ساختار ویژگی‌ها تمرکز دارد. علاوه بر این، برای در نظر گرفتن توزیع غیریکنواخت محلی اعوجاج در تصاویر، یک لایه کاملاً کانولوشن مبتنی بر ناحیه برای استفاده از اطلاعات در اطراف مرکز تصویر ورودی اضافه شده است.

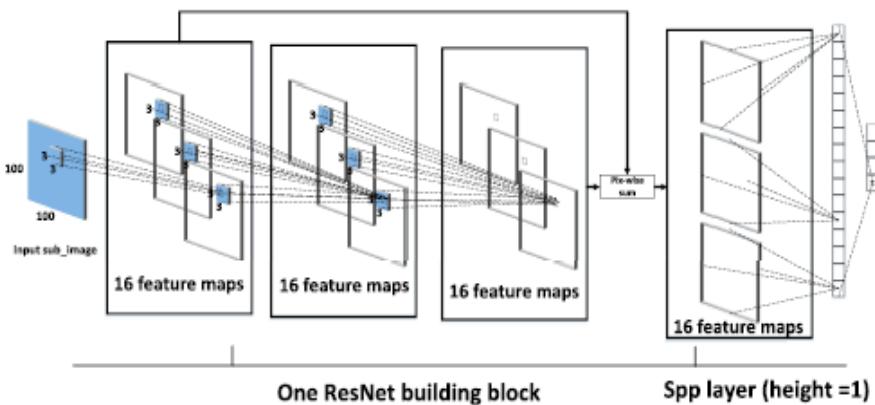
کانگ و همکاران [۱۱] از پیشگامان استفاده از شبکه عصبی کانولوشنال در روش NR-IQA هستند. مدل کانگ و همکاران از یک لایه کانولوشن با ادغام بیشینه و کمینه، دو لایه کاملاً متصل و یک گره خروجی تشکیل شده است. در ساختار شبکه، یادگیری ویژگی و رگرسیون در یک فرآیند بهینه‌سازی ادغام می‌شوند که منجر به مدل مؤثرتری برای تخمین کیفیت تصویر می‌شود. با این وجود آن‌ها از مجموعه فشرده‌ای از فیلترهای خطی استفاده کردند که قادر به استخراج ویژگی‌های پیچیده غیرخطی برای ارزیابی کیفیت تصویر بدون مرجع نیستند.

از طریق رویکرد NR-IQA مبتنی بر شبکه عصبی عمیق، ویژگی‌های سطح پایین و بالا موجود در تصویر به‌طور خودکار در طول آموزش شبکه عصبی عمیق استخراج می‌شوند. به این ترتیب در واحد کانولوشن، ویژگی‌ها توسط شبکه عصبی کانولوشنال استخراج می‌شوند و در واحد رگرسیون میزان کیفیت تصویر پیش‌بینی می‌شود که این پیش‌بینی می‌تواند از نوع سطح کیفیت یک تصویر و یا اینکه یک مقدار عددی به عنوان میزان کیفیت تصویر باشد.

لی و همکاران [۱۲] رویکردی را برای ارزیابی کیفیت تصویر با یک شبکه عصبی عمیق که شامل شبکه کانولوشن ResNet و یک لایه ادغام هرم فضایی است بر روی مجموعه داده LIVE، ارائه کردند. حاصل این پژوهش یک روش یادگیری انتها به انتهای انتها است به این صورت که تصویر به عنوان

<sup>۱</sup>Spatial Pyramid Pooling (SSP)

ورودی به سیستم داده می‌شود و به طور مستقیم میزان کیفیت نتیجه می‌شود. مدل لی و همکاران از نوع طبقه‌بندی است. لایه کاملاً متصل احتمال تعلق تصویر به سطح خاصی از کیفیت را نتیجه می‌دهد. شکل ۹-۲ نمایی از معماری شبکه ارائه شده توسط لی و همکاران را نشان می‌دهد. شبکه‌های عمیق کانولوشنال تصاویر با ابعاد یکسان را به عنوان ورودی می‌پذیرند. در مدل لی و همکاران، اضافه کردن لایه SSP قبل از لایه کاملاً متصل، شبکه کانولوشنال را قادر می‌سازد تصاویر با ابعاد دلخواه را دریافت کند. علاوه بر این لایه SPP شبکه را در برابر تغییر اشکال موارد درون تصویر تا حدی قوی می‌کند.



شکل ۹-۲ : معماری مدل لی و همکاران [۱۲]

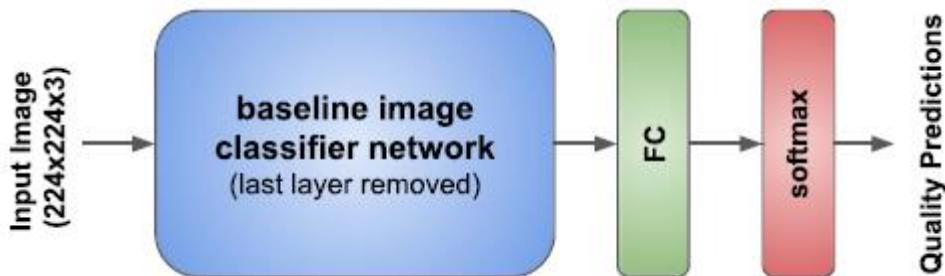
طالبی و میلانفر [۳۸] مدلی ارائه کردند که به جای پیش‌بینی میانگین امتیاز نظرها، توزیع نمرات کیفیت ادراکی کیفیت تصاویر را با استفاده از یک شبکه عصبی کانولوشنال پیش‌بینی می‌کند. مدل آن‌ها می‌تواند نه تنها برای امتیازدهی به کیفیت تصاویر به طور قابل اعتماد و با همبستگی بالا با درک انسان به کار رود، بلکه به منظور بهینه‌سازی الگوریتم‌های ویرایش و بهبود عکس در خط لوله عکاسی نیز استفاده شود.

طالبی و میلانفر از یک شبکه از پیش آموزش‌یافته بر روی مجموعه داده ImageNet در چارچوب خود استفاده کردند. آخرین لایه شبکه طبقه‌بندی کننده با یک لایه کاملاً متصل جایگزین شده است تا کیفیت تصویر را پیش‌بینی کند. وزن‌های لایه کاملاً متصل اضافه شده به طور تصادفی

مقداردهی اولیه شدند. مدل آنها در شکل ۲-۱۰ آمده است. تکنیک طالبی و میلانفر بر روی سه معماری شبکه عصبی کانولوشنال MobileNet [۳۹]، VGG16 [۴۰] و INCEPTION-V2 [۴۱] آموزش پیاده سازی شد. آنها مدل خود را توسط سه مجموعه دادگان AVA، LIVE و TID2013 و آموزش دادند. نتیجه ارزیابی مدل طالبی و میلانفر بر اساس معیار SROCC در جدول ۲-۲ آمده است.

جدول ۲-۲ : ارزیابی مدل طالبی و میلانفر بر اساس معیار SROCC

AVA	TID2013	LIVE	مجموعه دادگان آزمون	
			مجموعه دادگان آموزش	
۰/۲۰۰	۰/۳۲۷	۰/۶۳۷	LIVE	
۰/۰۸۷	۰/۷۵۰	۰/۱۵۵	TID2013	
۰/۶۱۲	۰/۴۳۲	۰/۵۴۳	AVA	



شکل ۲-۲ : مدل طالبی و میلانفر [۳۸]

### ارزیابی کیفیت تصویر متنی

ارزیابی کیفیت تصویر می‌تواند به صورت خاص بر روی تصاویر متنی صورت بگیرد. هدف ارزیابی کیفیت تصویر متنی ساخت یک مدل محاسباتی است که بتواند درجه تخریب تصاویر سند را پیش‌بینی کند. عملکرد بلاذرنگ سیستم‌های پردازش و تجزیه و تحلیل اسناد بر اساس نمرات کیفیت تخمین زده شده می‌تواند به حفظ، سازماندهی، شناسایی و بازیابی اطلاعات از تصاویر متنی کمک کند. در ادامه به چندین رویکرد این حوزه که در مطالعه ادبیات تحقیق در حوزه تصاویر متنی به آن برخورد شد، اشاره می‌شود.

پنگ و همکاران [۴۲] یک روش یادگیری نیمه نظارتی را برای استخراج ویژگی متن مبنی بر رویکرد یادگیری ماشین ارائه کردند که برای تخمین کیفیت تصاویر متنی از کارایی تشخیص کاراکتر نوری<sup>۱</sup> به عنوان معیار اندازه‌گیری استفاده می‌شود [۴۳]. معمولاً معیارهایی که برای ارزیابی کیفیت تصویر متنی ارائه شده‌اند از دقت OCR استفاده می‌کنند. این معیارها همیشه با درک انسان از کیفیت تصویر مرتبط نیستند. هنگام بهبود تصاویر متنی با هدف بهبود خوانایی در اسناد تاریخی که عملکرد OCR پایین است، درک انسان از کیفیت مهم است. هدف پژوهش [۴۴] طراحی سیستمی است که امکان یادگیری و تخمین درک انسان از کیفیت تصویر متنی را فراهم می‌کند به این صورت که کیفیت تصاویر متنی در سطح کاراکتر با استفاده از شبکه عصبی ارزیابی می‌شود. ویژگی‌های مختلف سطح کاراکتر بر اساس مورفولوژی ریاضی و مشخصه فضایی برای آموزش شبکه عصبی و ارزیابی کیفیت کاراکتر تصاویر استفاده شده است.

در پژوهش [۴۶] شبکه‌های عصبی کانولوشنال از پیش آموزش دیده برای ارزیابی کیفیت تصویر را با روش‌های ارزیابی با استفاده از ویژگی‌های استخراج شده از یک شبکه عصبی کانولوشنال به خوبی تنظیم شده مقایسه کرده و مدلی به نام DeepBIQ که کیفیت تصویر را با ادغام میانگین امتیازات پیش‌بینی شده در چندین ناحیه تصویر اصلی تخمین می‌زند، پیشنهاد می‌کند.

در مقاله [۴۷]، یک شبکه عصبی عمیق سه لایه مت Shank از رمزگذار، رمزگشا و پیش‌بینی برای شناسایی یک خط متنی و پیش‌بینی کیفیت آن ارائه شده است. کیفیت کلی تصویر متنی در نهایت با استفاده از میانگین وزن دار به عنوان استراتژی ادغام بیشینه بر روی کیفیت‌های به دست آمده برای خطوط متن محاسبه شده است.

یک مدل ارزیابی کیفیت تصویر متنی بدون تصویر مرجع بر اساس استخراج قسمت‌های غیر همپوشانی و محاسبه اختلاف ویژگی‌های گوسی از هر قسمت در پژوهش [۴۸] ارائه شده است. همچنین برای پیش‌بینی کیفیت قسمت‌های تصویر یک نگاشت کیفیت بر روی کلمات بصری ایجاد گردید. در نهایت برای پیش‌بینی کیفیت تصویر متنی ادغام میانگین آعمال گردید.

<sup>۱</sup>Optical Character Recognition (OCR)

<sup>۲</sup>Average Pooling

در مورد الگوریتم‌های ارزیابی بر اساس ویژگی‌هایی که توسط مدل یادگیری استخراج می‌شوند، لی کنگ و همکاران [۵۰] روشی بر اساس یادگیری عمیق برای بررسی کیفیت تصاویر اسناد با استفاده از یک مدل ۶ لایه شبکه عصبی کانولوشنی ارائه داده است. این روش به این صورت عمل می‌کند که در ابتدا تصویر متنی به قسمت‌هایی تقسیم می‌شود و قسمت‌هایی غیر متنی تصویر با استفاده از روش باینری کردن اوتسو<sup>۱</sup> جدا می‌شوند تا از سردرگمی در فرآیند آموزش یا آزمایش جلوگیری شود.

در مدل لی کنگ و همکاران ویژگی‌های توصیفی کیفیت تصویر به‌طور خودکار از قسمت‌های متنی تصویر با استفاده از دولایه کانولوشنال که به ترتیب از ۴۰ و ۸۰ فیلتر تشکیل شده‌اند استخراج می‌شوند. برای کاهش ابعاد ویژگی‌های استخراج شده یک لایه ادغام حداقل<sup>۲</sup> بعد از لایه‌های کانولوشنال قرار می‌گیرد.

مدل پیشنهادی لی کنگ و همکاران با دولایه کاملاً متصل با ۱۰۲۴ نورون که برای نگاشت ویژگی‌های استخراج شده با دقت تشخیص کاراکتر نوری آموزش دیده‌اند خاتمه می‌یابد. میزان کیفیت برای هر قسمت متنی انتخاب شده به دست می‌آید. کیفیت تصویر ورودی با میانگین‌گیری از کیفیت تمام این قسمت‌ها محاسبه می‌شود. با این حال روش آموزش در این روش به نمونه‌های برچسب‌گذاری شده عظیمی نیاز دارد که کمیاب و یا در دسترس نیستند. معماری شبکه ارائه شده در این پژوهش در شکل ۱۲-۲ آمده است.



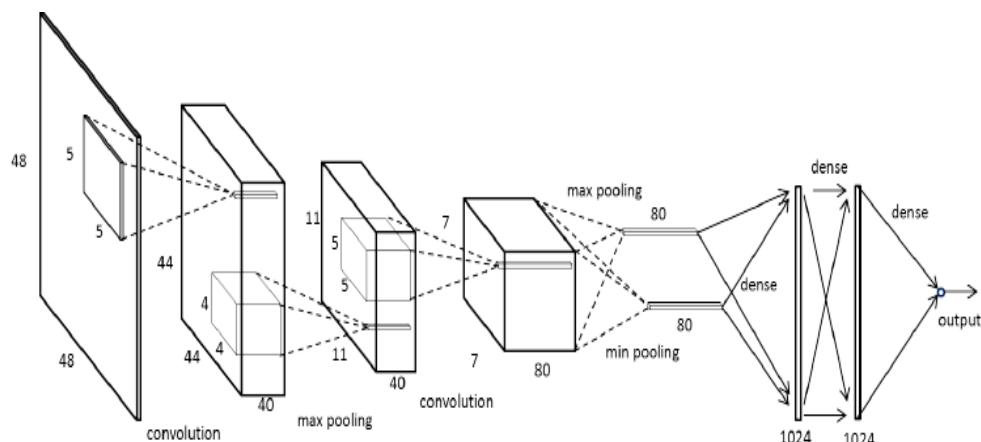
شکل ۱۱-۲ : مراحل پیش‌پردازش تصویر در مدل لی کنگ و همکاران [۵۰]

<sup>۱</sup>Otsu Binarization

<sup>۲</sup>MinMaxPooling

شکل ۱۱-۲ مراحل پیشپردازش تصویر ورودی مدل لی کنگ و همکاران را نشان می‌دهد. به ترتیب از چپ به راست، تصویر اصلی، تصویر نرمال شده محلی تصویر اصلی، تصویر باپنری شده به روش اوتسو و قسمت‌های متنی تصویر اصلی را نشان می‌دهد.

در پژوهش [۵۱] یک مدل برای ارزیابی کیفیت تصاویر متنی ارائه شده است که در آن شبکه عصبی کانولوشنال و شبکه عصبی بازگشتی یکپارچه شده‌اند تا یک فرایند عمل اجمالی تعاملی<sup>۱</sup> را تشکیل دهند که به‌طور خودکار توجه مدل را به قسمت‌های متنی یک سند جلب می‌کند. این قسمت‌های متنی به‌طور جمعی برای ارزیابی کیفیت تجزیه و تحلیل می‌شوند.



شکل ۱۲-۲ : معماری شبکه لی کنگ و همکاران [۵۰]

#### ۴-۲ جمع‌بندی

با مشاهده توانایی قدرتمند شبکه‌های عصبی عمیق در یادگیری مؤثر ویژگی‌های تصویر که توسط بینایی انسان استفاده می‌شود و اهمیت موضوع تحقیق در کاربردهای مختلف این پژوهش در صدد است گامی هرچند کوتاه در راستای بهبود عملکرد مدل ارزیاب کیفیت تصویر ایفا کند.

<sup>۱</sup>Glimpse Action

## **فصل ۳:**

**روش پیشنهادی**

در این فصل به شرح مسئله تحقیق می‌پردازد. در ادامه روش پیشنهادی مطرح می‌گردد و به روند پیاده‌سازی و آموزش آن می‌پردازد. درنهایت چگونگی ارزیابی عملکرد مدل پیشنهادی و نتیجه آزمایش آن بر روی دادگان مختلف بیان می‌شود.

### ۱-۳ تعریف دقیق مسئله

مسئله مورد نظر در این پژوهش ارزیابی خودکار کیفیت تصویر بدون تصویر مرجع است. با توجه به این‌که کاربر نهایی در برخی کاربردهای تصویر انسان است، هدف ارائه یک مدل مبتنی بر یادگیری عمیق است که بتواند کیفیت تصویر را نزدیک به میزان ارزیابی انسان در مورد کیفیت یک تصویر تخمین بزند.

یک سیستم تشخیص بیماری پوستی را در نظر بگیرید. در یک کلینیک درمانی این سیستم برای بررسی بیماری پوستی بیماران استفاده می‌شود. در مرحله آزمون پس از عکس‌برداری از ناحیه مورد نظر پوست فرد بیمار، سیستم پیش از آن‌که به تشخیص بیماری بر اساس تصویر تهیه شده اقدام کند، لازم است ابتدا از میزان کیفیت لازم تصویر ورودی اطمینان حاصل کند. در صورتی که تصویر ورودی از کیفیت کافی برخوردار نباشد فرایند تصویربرداری مجدد باید تکرار شود. اگر تصویر دارای انحرافاتی باشد و سیستم وارد فرایند تشخیص بیماری شود، ممکن است به درستی بیماری را تشخیص ندهد.

در سامانه قوه قضاییه لازم است مدارک و اسناد مربوطه بارگزاری شوند. قاضی با مطالعه این مدارک حکم صادر می‌کند. اگر این مدارک از کیفیت مناسبی برخوردار نباشند توسط قاضی قابل استفاده نیستند و فرایند رسیدگی متوقف می‌شود. بنابراین لازم است به صورت خودکار سامانه اجازه ارسال اینگونه تصاویر را ندهد تا فرایند رسیدگی به پرونده را سرعت بخشد. به همین دلیل ارزیابی خودکار کیفیت تصویر حائز اهمیت است و در هر حوزه کاربردی تصویر می‌تواند یک مرحله ضروری محسوب شود. در این پژوهش مدلی ارائه شده است که کیفیت یک تصویر را به طور خودکار ارزیابی می‌نماید.

## ۲-۳ روش حل مسئله

ارزیابی کیفیت یک تصویر به صورت خودکار مسئله‌ای است که این مطالعه در صدد ارائه راه حل برای آن انجام شده است. یک ناظر انسانی می‌تواند کیفیت تصاویر را در حوزه کاربردی ارزیابی کنند؛ اما فرایند زمان‌بری خواهد بود. از جهت دیگر خطای انسانی را نمی‌توان در این زمینه نادیده گرفت؛ بنابراین خودکارسازی فرایند ارزیابی کیفیت تصویر مورد توجه است.

از آنجا که ارزیابی کیفیت تصویر فرایندی پرکاربرد در به کارگیری هر سیستم مبتنی بر بینایی کامپیوتر محسوب می‌شود ارائه یک مدل معتبر و بهینه که کیفیت تصویر را با کمترین خطا ارزیابی کند ضروری است. برای شناسایی الگوهای ویژگی‌هایی که به راحتی قابل تشخیص نیستند، روش‌های مبنی بر هوش مصنوعی بسیار مناسب هستند. می‌توان از روش‌های مبنی بر هوش مصنوعی به عنوان ابزاری برای کشف شاخص‌هایی که بر میزان کیفیت تصویر موثر هستند استفاده شود.

در این پژوهش از روش یادگیری انتقالی برای ارائه یک مدل NR-IQA استفاده می‌شود.

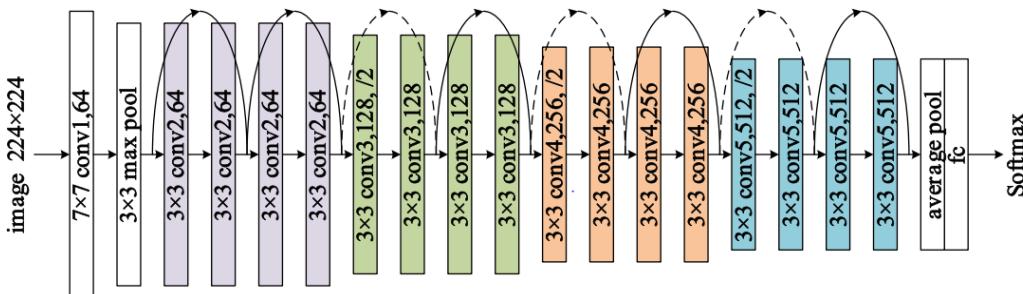
## ۳-۳ مدل پیشنهادی مبتنی بر یادگیری عمیق

رویکرد پیشنهادی ارائه یک مدل برای ارزیابی کیفیت تصویر مبتنی بر بازنظمیم یک شبکه عصبی کاملاً عمیق است. مدل پیشنهادی از دو بلوک اصلی تشکیل شده است که در بلوک اول آن از یک شبکه از پیش آموزش یافته برای استخراج ویژگی‌های تصویر استفاده می‌شود. انتخاب شبکه از پیش آموزش دیده در ارائه راه حل مبتنی بر یادگیری عمیق و به روش یادگیری انتقالی گام نخست و مهمی محسوب می‌شود. در این پژوهش از شبکه عصبی کانولوشنال ResNet18 که بر روی پایگاه داده عظیم ImageNet آموزش یافته است، استفاده گردید.

شبکه‌ی ResNet [۵۲] یک مدل قدرتمند است که در کارهای مربوط به حوزه‌ی بینایی کامپیوتر<sup>۱</sup> استفاده می‌شود. این شبکه با حل مشکل محوشدگی گرادیان و انفجار گرادیان به استفاده از شبکه‌های عمیق کمک بزرگی کرد. ResNet از معماری VGGNet از الهام گرفته شده است با این تفاوت که به این شبکه اتصالات میان بر اضافه شده است. منظور از اتصال میان بر این است که از یک

<sup>۱</sup>Computer Vision

یا چندلایه عبور می‌کند و آن‌ها را در نظر نمی‌گیرد؛ در واقع بهنوعی میانبر می‌زند و یک لایه را به لایه‌ی دورتر متصل می‌کند. شبکه‌ی ResNet در چندین نسخه که در تعداد لایه‌های شبکه در معماری‌شان تفاوت دارند، توسعه یافته است. در این تحقیق از نسخه ResNet18 به عنوان ستون فقرات معماری مدل پیشنهادی استفاده گردید. شبکه ResNet18 شامل ۱۶ لایه کانولوشنال است. معماری شبکه ResNet18 در شکل ۳-۱ آمده است.



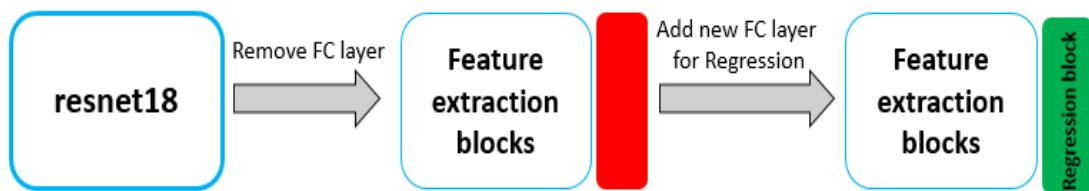
شکل ۳-۱ : معماری شبکه ResNet18

در بلوک دوم مدل پیشنهادی با توجه به آن که مسئله این تحقیق ارزیابی کیفیت تصویر است و یک مسئله رگرسیون محسوب می‌شود، یک بلوک رگرسیون قرار می‌گیرد. واحد رگرسیون با توجه به دانش حاصل از واحد کانولوشنال میزان کیفیت تصویر ورودی را خروجی می‌دهد. به منظور استفاده از شبکه از پیش آموزش داده ResNet18 برای مسئله رگرسیون ارزیابی کیفیت تصویر، ابتدا بلوک انتهایی آن که مربوط به طبقه‌بندی تصاویر ImageNet بود، از ساختار شبکه ResNet18 حذف گردید و یک بلوک رگرسیون جایگزین آن شد. بلوک رگرسیون ویژگی‌های استخراج شده توسط شبکه از پیش آموزش داده ResNet18 را به مقدار کمی کیفیت تصویر نگاشت می‌کند. شکل ۲-۳ نمایی از مراحل پیاده‌سازی مدل پیشنهادی را نشان می‌دهد.

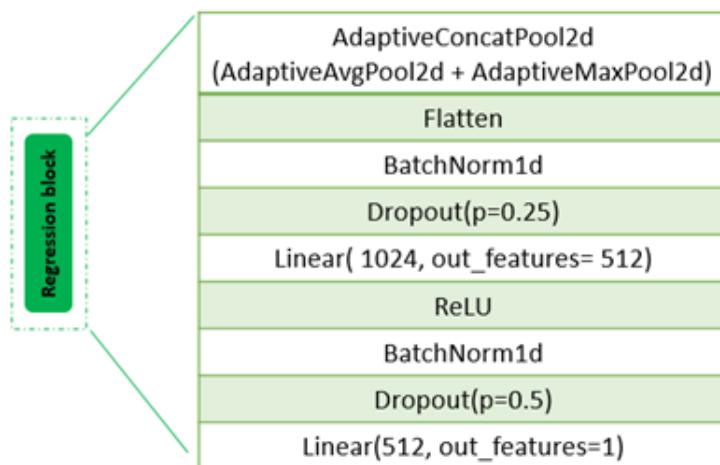
ساختار بلوک رگرسیون در شکل ۳-۳ آمده است. در ابتدا یک لایه ادغام ترکیبی شامل ادغام میانگین و ادغام بیشینه استفاده شده است. سپس یک لایه نرمال کننده دسته‌ای<sup>۱</sup> اعمال شد. به دلیل اینکه توزیع ورودی‌های هر لایه در طول آموزش متاثر از تغییر پارامترهای لایه‌های قبلی است،

<sup>۱</sup>Batch Normalization

آموزش شبکه‌های عصبی عمیق پیچیده است. اعمال نرخ‌های یادگیری پایین‌تر و مقداردهی دقیق اولیه پارامترها سرعت آموزش کاهش می‌یابد و آموزش مدل‌های پیچیده را سخت‌تر می‌کند. از طریق نرمال‌سازی ورودی‌های لایه‌های شبکه می‌توان این مشکل را حل کرد. تحقیقات نشان داده است که اعمال نرمال‌سازی در بخش‌هایی از معماری مدل و انجام نرمال‌سازی برای هر مینی دسته آموزشی در حل این مشکل نتیجه‌بخش بوده است [۵۴]. لایه نرمال‌سازی دسته‌ای با نرمال کردن توزیع داده امکان استفاده از نرخ‌های یادگیری بالاتری را فراهم می‌کند. استفاده از لایه‌ی نرمال‌سازی دسته‌ای باعث همگرایی سریع‌تر، ثابت فرایند آموزش مدل و بهبود عملکرد شبکه می‌شود



شکل ۲-۳ : نمایی از مراحل پیاده‌سازی مدل پیشنهادی



شکل ۳-۳ : بلوک رگرسیون مدل پیشنهادی

در لایه بعدی با هدف پیشگیری از بیش برآش مدل از لایه حذف تصادفی با نرخ ۲۵٪ استفاده شد. بعد از آن یک لایه خطی<sup>۱</sup> پیاده‌سازی شد. لایه خطی با استفاده از وزن‌ها و بایاس‌ها، تبدیل

<sup>۱</sup>Linear Layer

خطی را برای داده‌های ورودی اعمال می‌کند. برای پیاده‌سازی این تبدیل خطی به دو پارامتر تعداد ویژگی‌های ورودی و خروجی نیاز است: در معماری مدل پیشنهادی همان‌طور که در شکل ۳-۳ مشخص است، ۱۰۲۴ ویژگی ورودی و ۵۱۲ ویژگی خروجی به عنوان پارامترهای این لایه خطی تنظیم شدند.

برای اینکه یک نورون عصبی قابلیت غیرخطی داشته باشد لازم است یکتابع فعال‌ساز به ساختار آن اضافه شود. توابع فعال‌ساز متعددی وجود دارند از جمله سیگموید، تانژانت هایپربولیک و یکسوساز خطی. به منظور انتخاب تابع فعال‌ساز مناسب در این زمینه بررسی‌هایی صورت گرفت. یکی از نقاط ضعف تابع فعال‌ساز سیگموید، شیب ناپایدار است. به بیان دیگر، تابع سیگموید در کرانه بالا و پایین مسطح است. در طول بازگشت از طریق شبکه، با فعال شدن سیگموید، مشتق در گرهایی که خروجی آن‌ها در کرانه بالا و پایین صفر خواهد بود. این گرهای اشباع نامیده می‌شوند، وزنشان به روزرسانی نمی‌شود و فرایند یادگیری می‌شوند.

تابع تانژانت هایپربولیک هرچند برخلاف سیگموید، صفر محور هستند اما همانند سیگموید کران‌دار و همواره صعودی است و ضعف شیب ناپایدار را دارد. تابع یکسوساز خطی در برابر مسئله شیب ناپایدار، حداقل در ناحیه مثبت مقاوم است. این تابع به لحاظ محاسباتی هزینه کم‌تری دارد و بدین سبب از آن در یادگیری عمیق استفاده می‌شود.

بعد از تابع فعال‌ساز RELU یک لایه نرمال‌سازی دسته‌ای اضافه شد. در لایه بعدی نیز از لایه حذف تصادفی با نرخ ۵۰٪ استفاده شد. در آخر یک لایه خطی قرار داده شد. برای پارامترهای تبدیل خطی مقدار ۵۱۲ به عنوان تعداد ویژگی‌های ورودی و یک خروجی تنظیم جهت پیش‌بینی کیفیت تصویر تنظیم گردید.

#### ۴-۳ روش آموزش مدل پیشنهادی

محتوای یک تصویر نسبت به الگوهای ابتدایی و عمومی هر تصویر مثل لبه‌ها و خطوط در کیفیت ادراکی یک تصویر موثرتر است [۱۰]. لایه‌های ابتدایی شبکه عصبی ویژگی‌های سطح پایین و عمومی را استخراج می‌کنند و در لایه‌های نهایی ویژگی‌های سطح بالا را شناسایی می‌کنند. از این رو برای

مسئله حاضر که کیفیت تصویر با توجه به ادراک انسان باید ارزیابی شود، لازم است آموزش مدل بر روی لایه‌های نهایی متمرکز شود.

به منظور یافتن روش بهینه آموزش مدل، عملکرد آن پس از دو مرحله آموزش به روش‌های متفاوت سنجیده شد. در مرحله نخست فقط بلوک رگرسیون توسط داده‌های آموزشی و طی ۲۰ دوره<sup>۱</sup> آموزش داده شد و بلوک شبکه از پیش آموزش‌دیده ResNet18 قفل‌آشده تا پارامترهای آن طی فرایند پس انتشار به روزرسانی نشوند و بلوک رگرسیون بتواند الگوهای را از لایه‌های کانولوشنال آموخته‌شده در شبکه یاد بگیرد و از دانش آن استفاده کند و برای مسئله ارزیابی کیفیت تصویر به کار بگیرد. نتیجه عملکرد مدل بر روی دادگان اعتبارسنج به RMSE برابر با ۵۷۸/۰ منجر شد.

در مرحله دوم آموزش، بلوک ماقبل بلوک رگرسیون از حالت انجماد خارج شد و ۱ دوره مدل آموزش داده شد تا پارامترهای بلوک رگرسیون و بلوک ماقبل آن به روزرسانی شوند. سپس برای تکمیل آموزش و تنظیم بهتر پارامترها سر شبکه از حالت انجماد خارج گردید و وزن‌های کل مدل پیشنهادی طی ۲ دوره بازنظمی شدند. انجام مرحله دوم آموزش بهبود اندکی در عملکرد مدل ایجاد کرد و به RMSE برابر با ۵۸۸/۰ منجر شد. آموزش مدل پیشنهادی در مرحله دوم حتی با افزایش تعداد دوره‌های آموزش، پیشرفت شایانی نداشت و حتی منجر به بیش برازش مدل گردید. درنتیجه به دلیل هزینه‌بر بودن فرایند آموزش مدل، از این مرحله چشم‌پوشی شد و به آموزش مدل در مرحله اول اکتفا گردید.

در ادامه ابر پارامترها و پارامترهای مدل پیشنهادی بررسی می‌شوند.

### ۳-۵ ابر پارامترهای مدل<sup>۲</sup>

ابر پارامتر پارامتری است که مقدار آن، پیش از آغاز فرآیند یادگیری تعیین می‌شود.

#### ۳-۵-۱ نرخ یادگیری

ابر پارامتر نرخ یادگیری، ضریب تغییر وزن‌ها در روند آموزش شبکه را تعیین می‌کند. انتخاب

<sup>1</sup>Epoch

<sup>2</sup>Freeze

<sup>3</sup>Hyper Parameter

نرخ یادگیری مناسب بر کارایی مدل یادگیری عمیق موثر است.

اگر نرخ یادگیری کم باشد، با وجود همگرایی، زمان یادگیری زیادتر می‌شود و ممکن است بیش برآش رخ دهد، زیرا در فرایند آموزش هر بار که کل داده‌ها بررسی می‌شوند، این فرصت در اختیار مدل قرار می‌گیرد تا ویژگی‌های هر داده را با جزئیات به خاطر بسپارد. در مقابل مقدار بالای نرخ یادگیری باعث همگرایی سریع شبکه می‌شود؛ اما ممکن است از راه حل بهینه عبور کرده و واگرا شود.

در سال ۲۰۱۵ میلادی لزلی اسمیت [۵۵] ایده‌ای جدید برای تعیین نرخ یادگیری مناسب ارائه کرد. ایده او این بود که ابتدا با نرخ یادگیری کم آموزش شبکه شروع می‌شود. در هر مرحله از آموزش، نرخ یادگیری با ضریب ۲ افزایش می‌یابد. این عمل تا واگرایی شبکه تکرار می‌شود. در نهایت ۰/۹ آخرین نرخ یادگیری به عنوان نرخ یادگیری مناسب تعیین می‌شود.

کتابخانه Fastai [۵۶] این قابلیت را دارد که نرخ یادگیری مناسب را برای شبکه عصبی به روش لزلی اسمیت و به صورت خودکار تعیین کند. در این پژوهش از طریق این قابلیت Fastai، نرخ یادگیری مناسب ۰/۰۰۲ تعیین گردید.

### ۲-۵-۳ تعداد مراحل آموزش

در انتخاب تعداد مراحل آموزش اولین رویکرد باید این باشد که ابتدا یک تعداد دلخواه برای مراحل آموزش انتخاب شود و مدل را به آن میزان آموزش داد. سپس به نمودارهای زیان داده‌های آموزشی و اعتبارسنجی توجه کرد. اگر مدل حتی در دوره‌های پایانی آموزش همچنان در حال بهتر شدن است، مشخص می‌شود که برای مدت مناسب آموزش ندیده است و لازم است تعداد مراحل آموزش را افزایش داد.

در حالت دیگر اگر نمودار زیان برای داده‌های آموزش روند نزولی و برای داده‌های اعتبارسنج روند صعودی داشته باشد نتیجه‌گیری می‌شود که مدل به تعداد دفعه‌های زیادی آموزش دیده است و تنها مطابق ویژگی‌های دادگان آموزشی عمل می‌کند و برای دادگان جدید قدرت تشخیص ندارد. به بیان دیگر در این حالت بیش برآش رخ داده است و باید تعداد مراحل آموزش را کاهش داد.

### ۳-۵-۳ تابع فعال ساز

در شبکه عصبی برای اینکه یک نورون قابلیت غیرخطی داشته باشد یک تابع فعال ساز به ساختار آن اضافه می‌شود. نوع تابع ریاضی غیرخطی برای نورون‌های یک لایه، با ابر پارامتر نوع تابع فعال ساز مشخص می‌شود. توابع فعال‌سازی برای فعال شدن هر نورون در خروجی یک لایه تصمیم‌گیری می‌کنند.

### ۳-۵-۴ الگوریتم بهینه‌سازی<sup>۱</sup>

الگوریتم بهینه‌سازی چگونگی بهروزرسانی پارامترهای مدل در طول آموزش شبکه را مشخص می‌شود. بهینه‌ساز نقش مهمی در تعیین سرعت همگرایی شبکه به یک راه حل بهینه و ثبات آن دارد. الگوریتم بهینه‌ساز نه تنها برای تسريع در فرایند یادگیری و بهروزرسانی کنترل شده وزن‌ها مهم است، بلکه برای جلوگیری از بیش برازش نیز اهمیت دارد. در این تحقیق از بهینه‌ساز Adam استفاده شد.

بهینه‌ساز Adam ویژگی‌های هر دو بهینه‌ساز انتشار ریشه میانگین مرربع و تکانه<sup>۲</sup> را ترکیب می‌کند. بهینه‌ساز RMSprop ، با تنظیم نرخ یادگیری برای هر پارامتر بر اساس میانگین مربعات گرادیان‌های اخیر، از افت شدید و دائمی نرخ یادگیری در طول زمان جلوگیری می‌کند. این کار با محاسبه میانگین مربعات گرادیان‌ها و تقسیم نرخ یادگیری بر ریشه دوم این میانگین انجام می‌شود. درواقع این میانگین‌گیری کمک می‌کند تا بهروزرسانی وزن‌ها در مرحله‌های پایانی نیز ادامه یابد و بهبیان دیگر نرخ یادگیری را در حین بهینه‌سازی تطبیق دهد. این روش در شبکه‌های عصبی که با داده‌های پراکنده و ویژگی‌های متنوع سروکار دارند مؤثر است.

روش تکانه با در نظر گرفتن گرادیان‌های گذشته و حال، مسیر بهروزرسانی وزن‌ها را تسهیل می‌بخشد. با این کار تنظیمات وزنی پایدارتر و سریع‌تر به سمت کمینه سراسری حرکت می‌کنند. بهبیان دیگر در این بهینه‌ساز برای بهروزرسانی وزن‌ها علاوه بر گرادیان دسته فعلی به مقدار گرادیان

<sup>1</sup>Optimizer

<sup>2</sup>Root Mean Square Propagation (RMSprop)

<sup>3</sup>Momentum

دسته‌های قبل نیز توجه می‌کند تا فرایнд همگرایی به بهینه سراسری، بهصورت هموار اتفاق بیفت. این الگوریتم با استفاده از تاریخچه گرادیان‌ها، به حرکت وزن‌ها شتاب بیشتری می‌بخشد و به آن‌ها کمک می‌کند از کمینه‌های محلی عبور کنند؛ بنابراین با کاهش نوسان‌ها و جلوگیری از گیر افتادن در کمینه‌های محلی، کارایی را افزایش می‌دهد و بهینه‌ساز Adam با ترکیب دو معیار تغییر نرخ یادگیری بهصورت پویا و هموار ساختن فرایند همگرایی، قابلیت بهبود سرعت و پایداری در فرایند یادگیری را فراهم می‌کند.

### ۵-۵ تابع هزینه

انتخاب نوع اندازه‌گیری اختلاف بین پیش‌بینی‌های خروجی مدل و مقادیر واقعی را بر عهده‌ی تابع هزینه است. تابع هزینه، معیاری برای سنجش مدل از نظر توانایی در پیش‌بینی داده‌های جدید است. شبکه همواره سعی می‌کند که این مقدار را به حداقل برساند. از آنجا که یک ارزیابی کیفیت تصویر مسئله رگرسیون است و مدل باید یک مقدار عددی را پیش‌بینی کند تابع هزینه در این تحقیق خطای قدر مطلق مقادیر پیش‌بینی مدل و مقدار واقعی است.

### ۵-۶ اندازه‌ی دسته<sup>۱</sup>

به تعداد داده‌هایی که در طول به‌طور همزمان پردازش می‌شوند، اندازه‌ی دسته گفته می‌شود. تعداد بالای آن می‌تواند منجر به تسریع عملکرد شبکه و بهینه‌سازی منابع محاسباتی شود، اما ممکن است به حافظه‌ی بیشتری نیاز داشته باشد.

### ۷-۵-۳ نرخ حذف تصادفی

حذف تصادفی تکنیکی است که برای جلوگیری از بیش برآذش شبکه و تشویق آن به یادگیری ویژگی‌های مستقل و قوی استفاده می‌شود. نرخ حذف تصادفی، احتمال غیرفعال کردن تصادفی بخش خاصی از خروجی یک لایه را در طول آموزش شبکه مشخص می‌کند.

### ۸-۵-۳ لایه ادغام

نوع فیلتری که در لایه‌ی ادغام بر روی ورودی اعمال می‌شود، توسط این ابر پارامتر مشخص می‌شود.

<sup>۱</sup>Batch Size

پرکاربردترین انواع آن عبارت اند از:

• ادغام حداکثری: این نوع ادغام بر ویژگی‌های برجسته‌ی ورودی تمرکز و تأکید دارد.

• ادغام میانگین: این نوع ادغام بر ویژگی‌های کلی ورودی و حذف نویز تأکید دارد.

به طور کلی لایه ادغام ویژگی‌های موجود در هر ناحیه از نقشه‌ی ویژگی را که لایه‌ی کانولوشن تولید کرده است خلاصه می‌کند؛ درواقع مهم‌ترین ویژگی‌ها را انتخاب می‌کند و به مرحله‌ی بعد منتقل می‌کند. این باعث می‌شود مدل ویژگی‌های عمومی که مربوط به کیفیت تصویر است را یاد بگیرد و بر ویژگی‌های خاص متمرکز نشود و درنتیجه بتواند به عملکرد درستی داشته باشد. در ساختار بلوک رگرسیون مدل پیشنهادی از ادغام ترکیبی، استفاده گردید. به این صورت که یک مرحله ادغام میانگین و یک مرحله ادغام حداکثری اعمال شود. مقادیر تعیین شده برای ابر پارامترهای این پژوهش در جدول ۱-۳ گزارش شده است.

### ۶-۳ پارامترهای مدل

در شبکه‌های عصبی پارامترهای مدل همان وزن‌های معماري مدل هستند که برخلاف ابر پارامترها، با توجه به مسئله حین فرآیند یادگیری تعیین می‌شوند. مدل پیشنهادی مبتنی بر یادگیری انتقالی از دو بلوک مجزا تشکیل شده است که تعداد پارامترهای هر یک در جدول ۲-۳ زیر آمده است.

جدول ۱-۳ : مقادیر ابر پارامترهای مدل پیشنهادی

مقدار	ابر پارامترها
۰/۰۰۲	نرخ یادگیری
۳۲	اندازه دسته
۲۰	تعداد مراحل آموزش
ادغام میانگین ادغام حداکثر	لایه ادغام
دو لایه حذف تصادفی با نرخ ۰/۲۵ و ۰/۵	نرخ حذف تصادفی
Adam	الگوریتم بهینه‌سازی
RELU	تابع فعال‌ساز
خطای قدر مطلق	تابع هزینه

جدول ۲-۳: تعداد پارامترهای مدل پیشنهادی

تعداد پارامترهای واحد کانولوشنال	۱۱/۷۰۴/۳۸۴
تعداد پارامترهای واحد رگرسیون	۵۳۷/۴۷۲

## **فصل ۴:**

**نتایج، تجزیه و تحلیل و تفسیر آن‌ها**

این فصل ابتدا مجموعه دادگان مورد استفاده در حوزه ارزیابی کیفیت تصویر بررسی می‌شوند. سپس بستر پیاده‌سازی مدل پیشنهادی و نحوه ارزیابی مدل پیشنهادی مطرح می‌شود. در انتها نتایج حاصل از آزمایش‌های انجام شده و در نهایت کیفیت عملکرد مدل پیشنهادی بر روی داده‌های جدید بیان می‌گردد.

#### ۱-۴ مجموعه دادگان

نظرسنجی از تعدادی از کاربران انسانی در شرایط یکسان و استاندارد، یکی از راه‌های ارزیابی کیفیت تصویر، است. می‌توان برای سنجش یا توسعه معیارهای ارزیابی خودکار از امتیازهای حاصل از این نظرسنجی استفاده شود. فرایند تشکیل مجموعه دادگان موجود در حوزه ارزیابی کیفیت تصویر به این نحو است که مجموعه‌ای از تصاویر که تحت تاثیر مجموعه‌ای از تحریف‌های با شدت مختلف قرار گرفته‌اند، توسط مجموعه‌ای از کاربران امتیازدهی می‌شوند. امتیازهای حاصل نرمال‌سازی شده و پس از میانگین‌گیری به عنوان امتیاز هر تصویر ثبت می‌شود.

میانگین امتیاز نظرات<sup>۱</sup> ساده‌ترین معیار ارزیابی کیفی تصویر است. تعداد افراد زیادی پس از مشاهده یک تصویر نمره‌ای به عنوان میزان کیفیت آن، به تصویر منتب می‌کنند. اغلب در تعیین معیار MOS اصل تصویر در دسترس نیست و نمره ثبت‌شده همان برداشت کاربر از کیفیت تصویر بدون مقایسه آن با تصویر دیگر است.

معیار میانگین امتیاز نظرات تمایزی<sup>۲</sup> معیاری است که اصل تصویر و تصویر اعوجاج یافته هر دو در اختیار کاربر قرار می‌گیرند و کاربر بر اساس میزان شباهت آن دو تصویر امتیازی برای تصویر اعوجاج یافته ثبت می‌کند. این امتیاز در بازه [۰-۱۰۰] است که صد به معنای تفاوت کامل و صفر به معنای عدم وجود تفاوت آن دو تصویر است. هر چه مقدار عددی معیار DMOS بیشتر باشد، نشان‌دهنده افت کیفیت تصویر است. برخی از پایگاه‌های تصویر MOS، برخی DMOS و برخی نیز هر دو نوع معیار را به عنوان امتیاز تصاویر گزارش می‌دهند. در ادامه شرح مختصری از مجموعه دادگان موجود در حوزه ارزیابی تصویر آمده است.

<sup>۱</sup>Mean Opinion Scores (MOS)

<sup>۲</sup>Differential Mean Opinion Scores (DMOS)

#### A57 ۱-۱-۴ مجموعه داده

مجموعه داده A57 [۵۷] در دانشگاه کرنل آمریکا توسعه یافته است و شامل سه تصویر طبیعی مرجع-۸ بیتی در مقیاس خاکستری با اندازه ۵۱۲ در ۵۱۲ پیکسل و با مقادیر پیکسل در محدوده [۰-۲۵۵] است. این سه تصویر با شش نوع اعوجاج در سه سطح مختلف تحریف شدند و درنتیجه مجموعه داده A57 با ۵۴ تصویر اعوجاج یافته تشکیل شد. امتیاز ثبت شده برای هر تصویر توسط هفت ناظر تعیین شده است. مقیاس کیفیت نشان داده شده به ناظر انسانی برای رتبه بندی در بازه [۰-۱] و از نوع DMOS است. شکل ۱-۴ نه نمونه تصویر مجموعه داده A57 که به ترتیب از راست به چپ اعوجاج های ماتی گوسی، فشرده سازی JPEG و فشرده سازی JPEG2000 بر آنها اعمال شده است را نشان می دهد.



شکل ۱-۴: نمونه تصویر اعوجاج یافته مجموعه داده A57 [۵۸]

### ۲-۱-۴ مجموعه داده LIVE

مجموعه داده LIVE [۵۹] در آزمایشگاه مهندسی تصویر و ویدئو دانشگاه تگزاس ایجاد شده است. این مجموعه داده حاوی ۲۹ تصویر مرجع و ۷۷۹ تصویر تحریف شده رنگی با فرمت BMP است. امتیاز ثبت شده برای هر تصویر از نوع DMOS و در بازه [۰-۱۰۰] است.



شکل ۲-۴ : نمونه تصاویر مجموعه داده LIVE [۶۰]

شکل ۲-۴ نمونه تصاویر مجموعه داده LIVE را نشان می دهد که شامل ۴ تصویر اصلی و ۵ نسخه اعوجاج یافته هر تصویر است. اعوجاج های اعمال شده به ترتیب از چپ به راست شبیه سازی کanal محسوسی رایلی، ماتی گوسی، فشرده سازی JPEG2000، فشرده سازی JPEG و نویز جمع شونده سفید گوسی می باشند.

### ۳-۱-۴ مجموعه داده CSIQ

مجموعه داده CSIQ [۶۱] در دانشگاه اوکلاهاما آمریکا ایجاد شده است. این مجموعه داده حاوی ۳۰ تصویر مرجع رنگی با فرمت PNG است. تصاویر مرجع پایگاه تصویر CSIQ در شکل ۳-۴ قابل مشاهده است. بر روی تصاویر مرجع ۶ نوع اعوجاج در ۵ سطح مختلف اعمال شد. در مجموع ۸۶۶ تصویر اعوجاج یافته در مجموعه داده CSIQ موجود است. ۳۵ نفر برای امتیازدهی به تصاویر CSIQ شرکت کردند. امتیاز ثبت شده برای هر تصویر از نوع DMOS و در بازه [۰-۱۰۰] است. به گونه ای که امتیاز یک نشان دهنده بدترین کیفیت و صفر نشان دهنده بهترین کیفیت است. لیست اعوجاج های

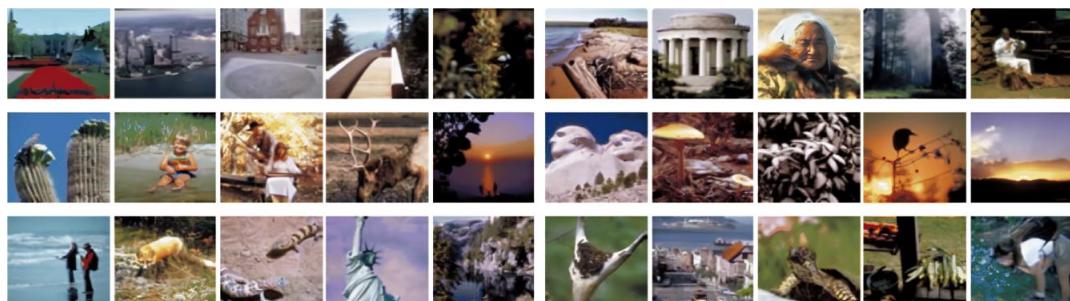
مجموعه دادگان A57 و LIVE و CSIQ در جدول ۱-۴ آمده است.

جدول ۱-۴ : اعوچاگهای مجموعه دادگان A57 LIVE و CSIQ

	معادل انگلیسی	تحریفها
A57 LIVE CSIQ	Additive Gaussian white noise	نویز جمع شونده سفید گوسی
A57	Quantization of the LH subbands of a 5-level DWT	کوانتیزاسیون زیر باندهای LH یک پنج سطحی DWT
	JPEG-2000 compression with the Dynamic Contrast-Based Quantization (DCQ) algorithm	فشرده سازی JPEG-2000 با الگوریتم کوانتیزاسیون مبتنی بر کنتراست پویا
NITS-IQA	Quantization	کوانتیزاسیون
	Pixelate Mosaic	موزاییک پیکسلی
	Motion Blur	تاری حرکت
	JPEG-XTcompression	فشرده سازی JPEG-XT
	Gaussian Noise	نویز گوسی
A57 LIVE NITS-IQA	JPEG-2000 compression	فشرده سازی JPEG-2000
	JPEG compression	فشرده سازی JPEG
A57 LIVE CSIQ NITS-IQA	Gaussian Blur	ماتی گوسی
LIVE	JPEG-2000 compression with bit error via Rayleigh fading channel simulation	فشرده سازی JPEG-2000 با خطای بیتی از طریق شبیه سازی کانال محوسازی رایلی
CSIQ NITS-IQA	Contrast change	تغییر کنتراست
CSIQ	Additive Gaussian pink noise	نویز جمع شونده صورتی گوسی

#### ۴-۱-۴ مجموعه دادگان TID2008 و TID2013

مجموعه داده TID2008 [۶۳] در دانشگاه صنعتی تمپرہ فنلاند ایجاد شده است. این مجموعه داده حاوی ۲۵ تصویر مرجع است که از مجموعه تصاویر رنگی شرکت کداک<sup>۱</sup> دریافت شده است. تمام این تصاویر از مناظر صحنه‌های طبیعی هستند به استثنای یک تصویر که ساختگی است. بر روی تصاویر مرجع این مجموعه داده تصویری، ۱۷ نوع اعوجاج مختلف و هر اعوجاج در چهار شدت متفاوت اعمال شده است. درمجموع ۱۷۰۰ تصویر رنگی در فرمت BMP و دارای اعوجاج در پایگاه تصویر TID2008 موجود است. برای برچسب‌گذاری میزان کیفیت تصاویر از سه کشور فنلاند، ایتالیا، اوکراین ۸۳۸ فرد در امتیازدهی به تصاویر مشارکت کردند.



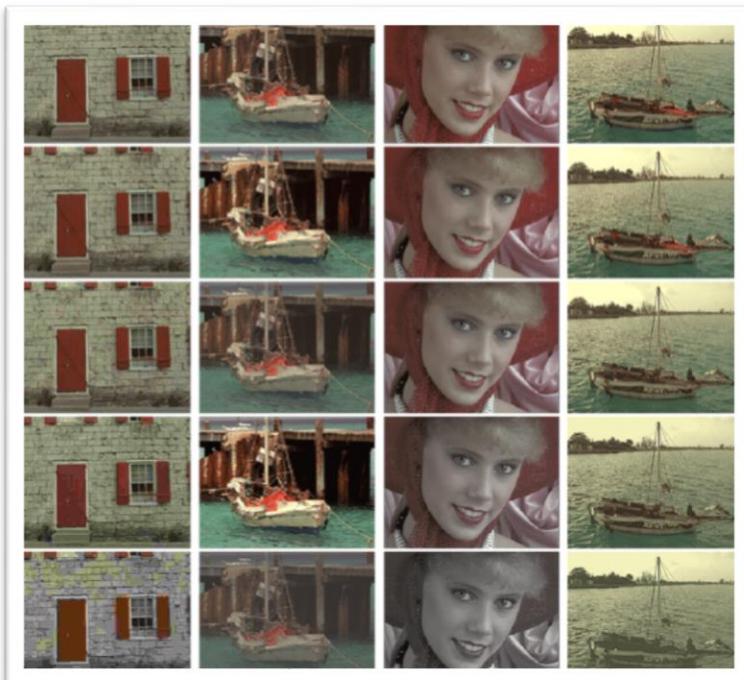
شکل ۳-۴ : ۳۰ تصویر مرجع مجموعه داده CSIQ [۶۲]



شکل ۴-۴ : تصاویر مرجع مجموعه داده TID2013 [۶۴]

<sup>۱</sup>Kodak

مجموعه داده TID2013 [۶۴] توسعه یافته مجموعه داده TID2008 است. ۷ اعوجاج و یک سطح اعوجاج جدید نسبت به مجموعه داده TID2008 به مجموعه داده TID2013 اضافه شده است. اعوجاج‌های اعمال شده بر روی مجموعه دادگان TID2013 و TID2008 در ۵-۱-۴ گزارش شده است. مجموعه داده تصویری TID2013 بر اساس میانگین نظر ۹۸۵ نفر از پنج کشور فنلاند، فرانسه، ایتالیا، اوکراین و ایالات متحده آمریکا برچسب‌گذاری شده است. درنتیجه در مجموعه داده TID2013 تصاویر تحریف شده با اعمال ۲۴ نوع اعوجاج در ۵ سطح اعوجاج متفاوت بر روی هر تصویر مرجع به دست می‌آیند که در مجموع شامل ۳۰۰۰ تصویر دارای اعوجاج و ۲۵ تصویر مرجع است. تصاویر مرجع مجموعه دادگان TID2013 و TID2008 در شکل ۴-۴ نشان داده شده است.



شکل ۴-۵ : نمایی از سطوح مختلف چهار اعوجاج مجموعه داده TID2013 [۶۴]

شکل ۴-۵ نمونه‌ای از پنج سطح مختلف چهار اعوجاج متفاوت را نشان می‌دهد. اعوجاج‌ها به ترتیب از چپ به راست عبارت هستند از فشرده‌سازی JPEG، تغییر کنترast، تغییر اشباع رنگ و کوانتیزاسیون رنگ تصویر با دیتر که از بالا به پایین شدت اعوجاج‌ها افزایش یافته است. ۲۵ نوع اعوجاج مجموعه داده TID2013 که بر روی یک تصویر اعمال شده است در شکل ۶-۴ آمده است.



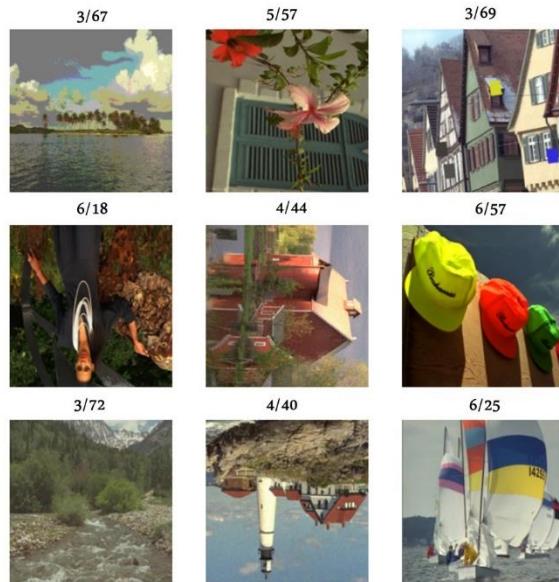
شکل ۴-۲۴: ۲۴ نوع اعوجاج مجموعه داده TID2013 بر روی یک تصویر

امتیازدهی به تصاویر TID2008, TID2013 با رویکرد نوینی صورت گرفت. طی ۹۸۵ آزمایش یک تصویر مرجع و دو تصویر اعوجاج یافته به طور همزمان به مخاطب نمایش داده شدند که نمایی از محیط نظرسنجی کیفی تصاویر TID2008, TID2013 در شکل ۷-۴ آمده است. در هر آزمایش، هر تصویر تحریف شده در ۹ مقایسه تصادفی زوجی شرکت داده می‌شود. ناظر از میان تصاویر تحریف شده، تصویر بهتر را انتخاب می‌کند. به تصویر انتخاب شده یک امتیاز تعلق می‌گیرد. در مجموع ۹۸۵ آزمایش انجام شد و میانگین امتیاز نظرهای آن‌ها به عنوان معیار کیفیت هر تصویر که در بازه [۰-۹] است تحت عنوان MOS ثبت گردید. نمونه‌هایی از تصاویر اعوجاج یافته پایگاه داده TID2013 با مقادیر MOS متناظر هر تصویر در شکل ۸-۴ آمده است.



شکل ۷-۴ : نمایی از نرم‌افزار نمره دهنده تصاویر TID2008 و TID2013 [۶۳]

به منظور آموزش و ارزیابی عملکرد مدل پیشنهادی از مجموعه داده‌ی TID2013 استفاده گردید. به صورت تصادفی ۲۰ درصد از دادگان به عنوان داده آزمایشی و ۸۰ درصد دادگان برای آموزش شبکه تفکیک شدند. از میان دادگان آموزشی ۲۰ درصد نیز به عنوان داده اعتبارسنجی شبکه در طول فرایند آموزش در نظر گرفته شدند. از هر ۲۴ نوع اعوجاج برای هر تصویر مجموعه داده ۱۲۵ نمونه موجود است که به صورت تصادفی ۱۰۰ تصویر برای مرحله آموزش و ۲۵ تصویر برای مرحله آزمون مدل پیشنهادی استفاده شد.



شکل ۸-۴ : نمونه‌ای از تصاویر TID2013 با مقادیر MOS متناظر

#### ۵-۱-۴ مجموعه دادگان CCID2014 و CID2013

مجموعه داده CID2013 [۶۵] شامل ۴۰۰ تصویر است که تغییر کنتراست یافته ۱۵ تصویر مرجع است. در شکل ۹-۴ چند نمونه از تصاویر اصلی مجموعه داده CID2013 قابل مشاهده است. منحنی‌های مقعر<sup>۱</sup> و محدب<sup>۲</sup> و منحنی‌های درجه سه،<sup>۳</sup> منحنی‌های لجستیک<sup>۴</sup> و شیفت شدت روشنایی<sup>۵</sup> بر روی تصاویر مرجع اعمال گردیده است و نسخه‌هایی تغییر کنتراست یافته از آن‌ها ایجاد شده است.

مجموعه داده CCID2014 [۶۶] نسخه توسعه یافته CID2013 است. هر دو مجموعه داده دارای یک نوع اعوجاج هستند. مجموعه داده CCID2014 شامل ۶۵۵ تصویر است که در ایجاد آن‌ها علاوه بر توابعی که برای مجموعه داده CID2013 اعمال شده‌اند، تبدیل گاما<sup>۶</sup> و توابع ترکیبی (شیفت شدت روشنایی پس از اعمال لجستیک)<sup>۷</sup> نیز استفاده شده است. امتیاز ثبت شده برای تصاویر مجموعه دادگان CCID2014 و CID2013 از نوع MOS و در بازه‌ی [۱-۵] هستند.

<sup>۱</sup>Concave

<sup>۲</sup>Convex

<sup>۳</sup>Cubic

<sup>۴</sup>Logistic

<sup>۵</sup>Mean Shift

<sup>۶</sup>Gamma Transfer

<sup>۷</sup>Compound Function (Mean Shift Followed by Logistic)

جدول ۲-۴ : لیست اعوجاج‌های اعمال شده بر روی مجموعه دادگان TID2008 و TID2013

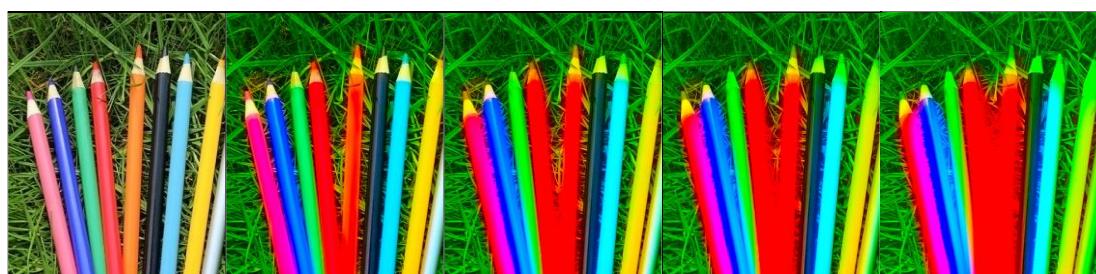
TID2013	TID2008	مخفف	معادل انگلیسی	نام اعوجاج
◆	◆	AGN	Additive Gaussian noise	نویز جمع شونده گوسی
◆	◆	ANC	Additive noise in components color	نویز جمع شونده در اجزای رنگی
◆	◆	SCN	Spatially correlated noise	نویز همبسته با مکان
◆	◆	MN	Masked noise	نویز ماسک شده
◆	◆	HFN	High freequently noise	نویز با فرکانس بالا
◆	◆	IN	Impulse noise	نویز نقطه‌ای
◆	◆	QN	Quantization noise	نویز کوانتیزاسیون
◆	◆	GB	Gussian blur	ماتی گوسی
◆	◆	DEN	Image denoising	ماتی حاصل از رفع نویز
◆	◆	JPEG	JPEG compression	فشرده‌سازی JPEG
◆	◆	JP2K	JPEG2000 compression	فشرده‌سازی JPEG2000
◆	◆	JGTE	JPEG transmission error	خطای انتقال JPEG
◆	◆	J2TE	JPEG2000 transmission error	خطای انتقال JPEG2000
◆	◆	NEPN	Non eccentricity pattern noise	نویز جابه‌جایی اندک بلوك
◆	◆	Block	Local block-wise distortions	اعوجاج بلوك محلی
◆	◆	MS	Mean shift (intensity shift)	تغییر میانگین روشنایی
◆	◆	CTC	change Contrast	تغییر کتراست
◆		CCS	Change of color saturation	تغییر اشباع رنگ
◆		MGN	Multiplicative Gaussian noise	نویز گوسی ضرب شونده
◆		CN	Comfort noise	نویز کامفورت
◆		ICNI	Lossy compression of noisy images	فشرده‌سازی با اتلاف تصاویر نویزی
◆		ICQD	Image color quantization with dither	کوانتیزاسیون تصویر رنگی با دیتر
◆		CHA	Chromatic aberrations	انحراف رنگی
◆		SSR	Sparse sampling and construction	نمونه‌برداری تنک و بازسازی تصویر



شکل ۹-۴ : نمونه‌هایی از تصاویر مرجع مجموعه داده CID2013 [۶۷]

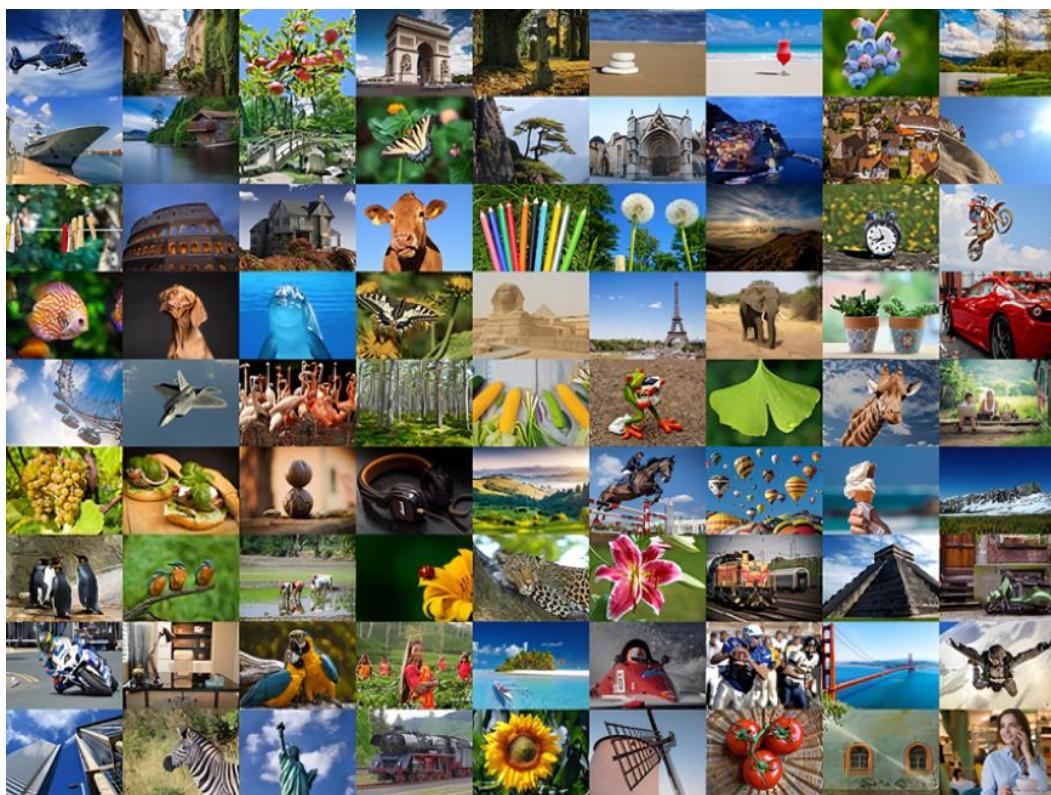
#### ۶-۱-۴ مجموعه دادگان KADID700K و KADID10K

مجموعه داده KADID10K [۶۸] شامل ۸۱ تصویر مرجع است. شکل ۱۱-۴، تصاویر اصلی مجموعه داده KADID10K را نشان می‌دهد. بر روی هر تصویر ۲۵ اعوجاج در ۵ سطح اعمال شده است. مجموعه داده KADID700K دارای ۱۴۰۰۰ تصویر اصلی با ۵ نسخه اعوجاج یافته است که در آن اعوجاج‌ها به‌طور تصادفی انتخاب شده‌اند. نتیجه اعمال اعوجاج انتشار رنگ بر روی یک تصویر در ۵ شدت مختلف در شکل ۱۰-۴ آمده است.



شکل ۱۰-۴ : اعوجاج انتشار رنگ در پنج سطح بر روی یک تصویر KADID10K

امتیاز ثبت‌شده برای هر تصویر در مجموعه داده KADID10K از نوع DMOS است. به این صورت که تصویر اعوجاج یافته و تصویر مرجع آن همزمان به کاربر نشان داده می‌شوند و کاربر با مقایسه این دو تصویر، از میان پنج توصیف کیفی مشخص شده برای میزان اعوجاج تصویر یک مورد را به عنوان کیفیت تصویر اعوجاج یافته ثبت می‌کند.

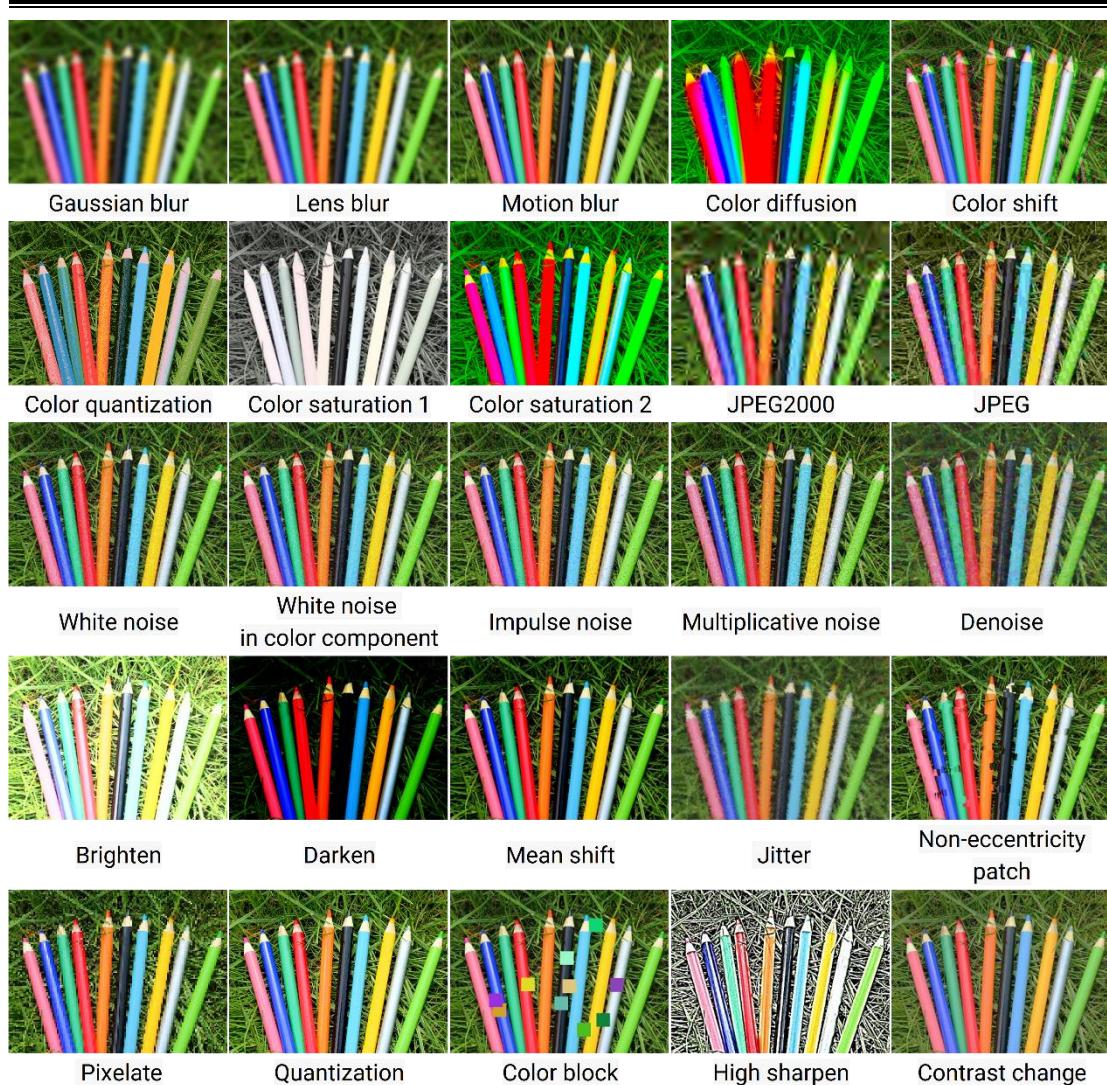


شکل ۱۱-۴ : تصاویر مرجع مجموعه دادگان KADID10K و KADID700K [۶۹]

اعوجاج‌های اعمال شده در دو مجموعه داده KADID10K و KADID700K را می‌توان به اعوجاج مربوط به تاری (گوسی، لنز، حرکت)، مربوط به رنگ (انتشار، جابجایی، کوانتیزاسیون، اشباع بیش از حد و اشباع زدایی)، فشرده‌سازی (JPEG JPEG2000)، نویز (سفید، رنگی، ضربه‌ای، ضربی، حذف نویز)، تغییر روشنایی (روشن، تاریک، جابجایی میانگین)، اعوجاج فضایی (جیتر، پیکسلی، کوانتیزاسیون)، وضوح و کنتراست گروه‌بندی کرد. شکل ۱۲-۴ نسخه اعوجاج یافته یک تصویر توسط ۲۵ اعوجاج در مجموعه دادگان KADID10K و KADID700K را نشان می‌دهد.

#### ۷-۱-۴ مجموعه داده AVA

مجموعه داده AVA [۷۰] شامل ۲۵۵۰۰۰ تصویر است که بر اساس ویژگی‌های زیبایی‌شناسی توسط عکاسان غیرحرفه‌ای رتبه‌بندی شده است. هر تصویر در فرایند ۹۰۰ چالش با موضوع‌های مختلف اخذ شده است. رتبه‌بندی تصویر در بازه [۱۰-۱۰۱] است که ۱۰ بالاترین امتیاز زیبایی‌شناسی مرتبط با یک تصویر است. شکل ۱۳-۴ نمونه‌ای از تصاویر مجموعه داده AVA که در چالشی با موضوع تصاویر منظره ثبت شده است، نشان می‌دهد.



شکل ۱۲-۴ : اعوچاگهای مجموعه دادگان KADID700K و KADID10K

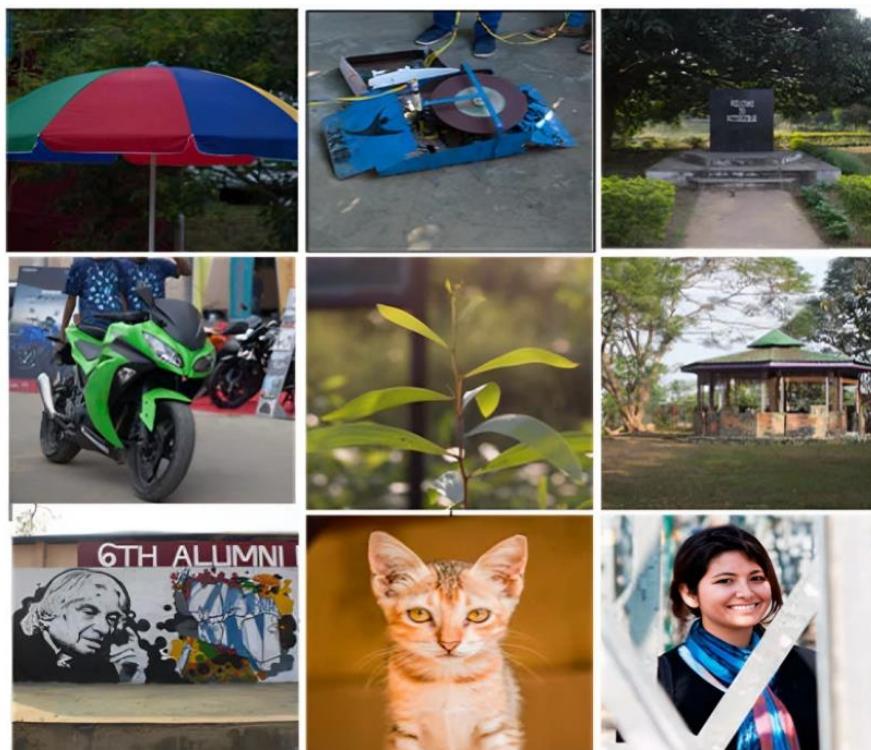


شکل ۱۳-۴ : نمونه تصاویر مجموعه داده AVA با موضوع چالش منظره [۳۸]

#### NITS-IQA ۸-۱-۴ مجموعه داده

مجموعه داده NITS-IQA [۷۱] شامل نه تصویر مرجع است که از محوطه موسسه ملی فناوری سیلچار<sup>۱</sup> اخذ شده‌اند. شکل ۱۴-۴ تصاویر اصلی مجموعه داده NITS-IQA را نشان می‌دهد. بر روی تصاویر مرجع، ۹ نوع اعوجاج مختلف که عمدتاً ناشی از فشرده‌سازی، نویز و مصنوعات تاری است، در پنج سطح از درجه پایین افت کیفیت تا درجه بالای افت کیفیت توسط برنامه فتوشاپ<sup>۲</sup> اعمال شده است. برای اعوجاج فشرده‌سازی از استانداردهای فشرده‌سازی JPEG و JPEG2000 و XT استفاده شده است. به‌طورکلی اعوجاج‌های این مجموعه داده همان اعوجاج‌هایی هستند که اغلب برای عکس‌های واقعی که توسط افراد با موبایل اخذ می‌شود، اتفاق می‌افتد. اعوجاج مجموعه داده NITS-IQA که بر روی ۹ تصویر اعمال شده است در شکل ۱۵-۴ آمده است.

در این بخش به بررسی مجموعه دادگان حوزه ارزیابی کیفیت تصویر پرداخته شد. جزئیات مجموعه دادگان پرکاربرد در جدول ۳-۴ آمده است.



شکل ۱۴-۴ : تصاویر مرجع مجموعه داده NITS-IQA [۷۱]

<sup>۱</sup>National Institute of Technology, Silchar

<sup>۲</sup>Adobe Photoshop CS6



[۷۱] انواع اعوجاج‌های مجموعه داده NITS-IQA

جدول ۳-۴: مجموعه دادگان حوزه ارزیابی کیفیت تصویر

بازه MOS/ DMOS	MOS/ DMOS	فرمت تصویر	نوع تصویر	تعداد مشاهده‌کنندگان	تعداد اعوجاج‌ها	تعداد تصاویر اعوجاج یافته	تعداد تصاویر مرجع	سال	مجموعه داده
[۰-۱۰۰]	DMOS	bmp	Color	۱۶۱	۵	۷۷۹	۲۹	۲۰۰۶	LIVE
[۰-۱]	DMOS	bmp	Gray	۷	۶	۵۴	۳	۲۰۰۷	A57
[۰-۹]	MOS	bmp	Color	۸۳۸	۱۷	۱۷۰۰	۲۵	۲۰۰۸	TID2008
[۰-۱]	DMOS	png	Color	۳۵	۶	۸۶۶	۳۰	۲۰۰۹	CSIQ
[۰-۹]	MOS	bmp	Color	۹۸۵	۲۴	۳۰۰۰	۲۵	۲۰۱۳	TID2013
[۱-۵]	DMOS	png	Color	۲۲۰۹	۲۵	۱۰۱۲۵	۸۱	۲۰۱۹	KADID10K

## ۲-۴ محیط پیاده‌سازی

و Keras از جمله کتابخانه‌های یادگیری عمیق هستند که برای استفاده در حوزه‌های مختلف قابل دسترس هستند. مدل‌های ارائه شده در این دو کتابخانه بر روی پایگاه داده عظیم ImageNet قابل دسترس هستند.

آموزش یافته‌اند. مزیت‌هایی که Fastai را از Keras و دیگر کتابخانه‌های یادگیری عمیق مجزا می‌کند در ذیل آمده است.

- بهینه‌شده از نظر پردازش گرافیکی
  - دارای یک نوع سیستم توزیع جدید برای پایتون به همراه سلسله‌مراتب معنایی برای تنسورها<sup>۱</sup>
  - دارای تابع بهینه‌سازی
  - دارای یک سیستم جدید که می‌تواند به هر بخشی از داده‌ها، مدل‌ها یا بهینه‌سازها دسترسی داشته باشد و آن را در هر نقطه از آموزش تغییر دهد.
- مدل‌هایی که در این پژوهش برای ارزیابی کیفیت تصویر مورداستفاده قرار گرفته است به زبان پایتون و در بستر Fastai پردازنده گرافیکی NVIDIA Tesla K80 و در محیط Google Colab پیاده‌سازی شده‌اند. Google Colab یک سرویس ابری رایگان توسط شرکت گوگل است. این سرویس امکان برنامه‌نویسی و امکان نصب و کار با کتابخانه‌های مطرح یادگیری عمیق را فراهم می‌آورد. با این سرویس ابری، می‌توان بدون نیاز به سخت‌افزار قدرتمند، روی یک پردازنده به صورت اینترنتی کار کرد. چراکه این امکان را فراهم می‌آورد تا از پردازنده‌های قدرتمند برای اجرای الگوریتم‌های هوش مصنوعی بر پایه پایتون به صورت رایگان استفاده شود.

### ۳-۴ بررسی عملکرد یک معیار

به منظور رصد عملکرد معیار ارزیابی کیفیت تصویر و بررسی این‌که چه میزان می‌تواند مقادیر MOS را به صورت دقیق پیش‌بینی کند، سه شاخص زیر بررسی می‌شوند.

#### ۱-۳-۴ دقت پیش‌بینی

میزان همبستگی بین مقادیر MOS و مقادیر پیش‌بینی شده برای کیفیت تصویر توسط مدل نشان‌دهنده دقت پیش‌بینی است. برای ارزیابی این همبستگی از ضریب همبستگی خطی پیرسون<sup>۲</sup> که در رابطه (۱-۴) تعریف شده است، استفاده می‌شود [۷۲]. در رابطه (۱-۴)  $n$  تعداد تصاویر،  $y_P$

<sup>۱</sup>Tensors

<sup>۲</sup>Pearson Linear Correlation Coefficient (PLCC)

مقدار پیش‌بینی شبکه،  $y_g$  مقدار MOS،  $\bar{y}_P$  میانگین مقادیر پیش‌بینی شبکه و  $\bar{y}_g$  میانگین مقادیر MOS را نشان می‌دهند.

$$PLCC = \frac{\sum_{i=1}^n (y_g - \bar{y}_g)(y_P - \bar{y}_P)}{\sqrt{\sum_{i=1}^n (y_g - \bar{y}_g)^2} \sqrt{\sum_{i=1}^n (y_P - \bar{y}_P)^2}} \quad (2-4)$$

می‌توان علاوه بر محاسبه PLCC، میزان خطای بین مقدار MOS و مقدار پیش‌بینی شبکه را به دست آورد و دقت پیش‌بینی را سنجید. برای محاسبه خطا می‌توان از ریشه میانگین مربعات خطای استفاده کرد که تعریف آن در رابطه (۳-۴) آمده است. در رابطه (۳-۴) منظور از  $y_p$  مقدار ارزیابی شده توسط شبکه،  $y_g$  مقدار MOS و  $n$  تعداد تصاویر است.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_p - y_g)^2} \quad (3-4)$$

### ۲-۳-۴ یکنواختی پیش‌بینی

یکنواختی پیش‌بینی، رابطه یکنواخت بین مقادیر پیش‌بینی شده را ارزیابی می‌کند. منظور از سنجش یکنواختی پیش‌بینی این است که چه میزان ترتیب مقادیر پیش‌بینی مدل با ترتیب مقادیر معیارهای کیفی مطابقت دارد. جهت بررسی این شاخص، از ضریب همبستگی اسپیرمن استفاده می‌شود.

$$SROCC = 1 - \frac{6}{N(N-1)} \sum_{i=1}^n \left( \left( RANK(y_{gi}) - RANK(y_{pi}) \right)^2 \right) \quad (4-4)$$

در رابطه (۴-۴) تعریف شاخص SROCC آمده است که در آن  $y_{pi}$  مقدار پیش‌بینی شبکه و  $y_{gi}$  مقدار MOS برای تصویر  $i$  ام و  $n$  تعداد تصاویر است [۷۲]. برای هر زوج  $(y_{gi}, y_{pi})$  یک زوج ترتیب  $RANK(y_{gi}), RANK(y_{pi})$  به این صورت تعریف

<sup>۱</sup>Root Mean Square Error (RMSE)

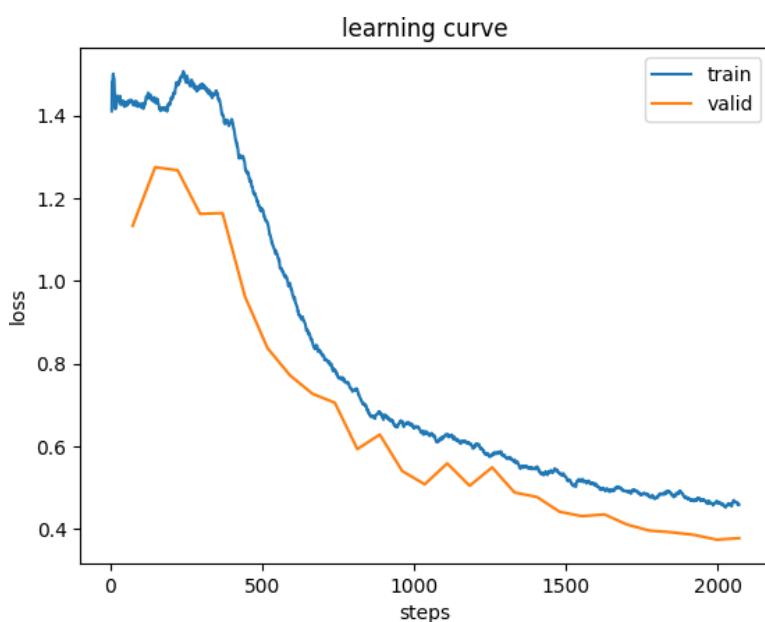
<sup>۲</sup>Spearmanrank-Order Correlation Coefficient (SROCC)

می‌شود که  $y_{gi}$  در لیست مرتب شده بر حسب مقادیر پیش‌بینی مدل در جایگاه  $k$ ام قرار دارد. درنتیجه در رابطه فوق ( $y_{gi}$ )  $RANK$  نشان‌دهنده جایگاه مقدار MOS تصویر نام در لیست مرتب شده بر اساس مقادیر پیش‌بینی مدل و ( $y_{pi}$ )  $RANK$  نشان‌دهنده جایگاه مقدار پیش‌بینی مدل تصویر نام در لیست مرتب شده بر اساس مقادیر MOS است.

برای هر دو شاخص، مقدار بیشتر نشان‌دهنده عملکرد بهتر معیار ارزیابی کیفیت تصویر است. می‌توان به جای MOS، معیار DMOS را در نظر گرفت.

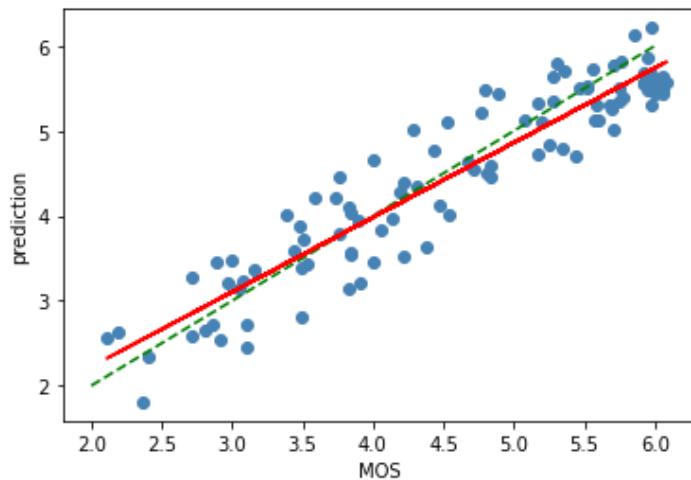
#### ۴-۴ نتایج حاصل از روش پیشنهادی

در این بخش نتایج حاصل از روش پیشنهادی آمده است. شکل ۱۶-۴ روند تغییر تابع هزینه برای داده‌های آموزش و اعتبارسنج در مرحله آموزش مدل پیشنهادی را نشان می‌دهد.



شکل ۱۶-۴ : نوسان تابع هزینه برای داده‌های آموزش و اعتبارسنج

شکل ۱۷-۴ مقادیر پیش‌بینی شده میزان کیفیت ۱۰۰ تصویر پایگاه داده TID2013 توسط مدل پیشنهادی که بر پایه شبکه ResNet18 است را نسبت به مقادیر MOS نشان می‌دهد. خط رگرسیون قرمز رنگ برازش پیش‌بینی مدل را نشان می‌دهد. میزان تطابق خط رگرسیون و خط سبزرنگ  $x = y$  نشان‌دهنده همبستگی بین پیش‌بینی مدل و مقدار MOS است.



شکل ۱۷-۴ : رگرسیون پیش‌بینی مدل پیشنهادی نسبت به MOS

عملکرد مدل پیشنهادی طی ۱۰ مرحله توسط سه معیار ارزیابی SROCC، RMSE و PLCC بررسی شد. هر مرحله از اجرای فرایند آموزش و ارزیابی مدل پیشنهادی ۷ دقیقه به طول انجامید. میانگین نتایج این ۱۰ مرحله در جدول ۴-۴ آمده است.

جدول ۴-۴ : نتایج ارزیابی مدل پیشنهادی

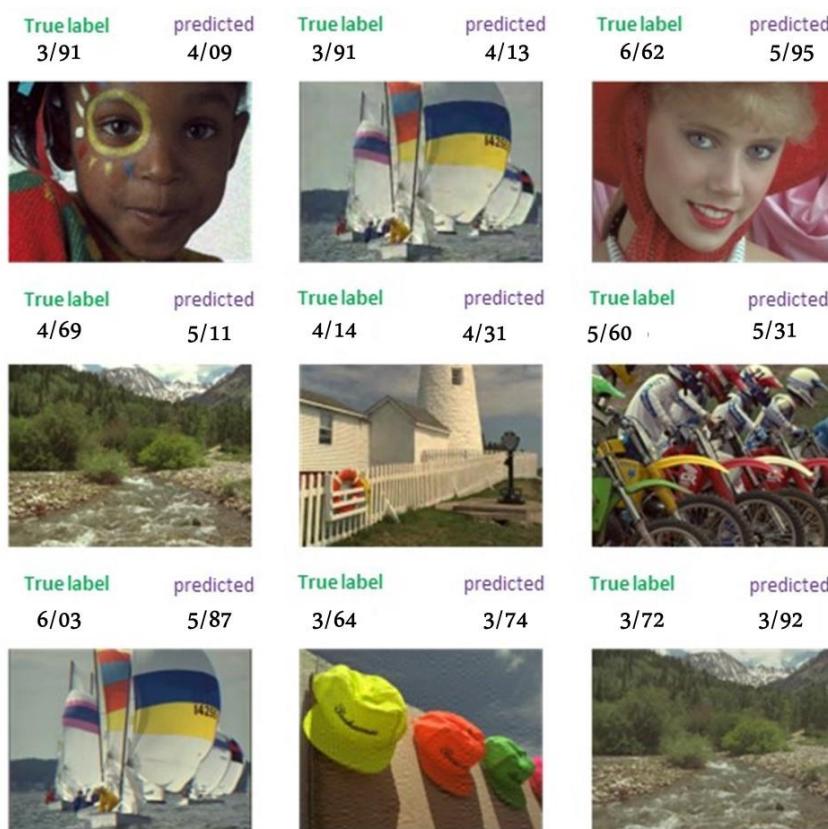
SROCC	PLCC	RMSE
۱	۰/۹۰	۰/۵۴

#### ۵-۴ بررسی عملکرد مدل آموزش‌دیده مبتنی بر روش پیشنهادی

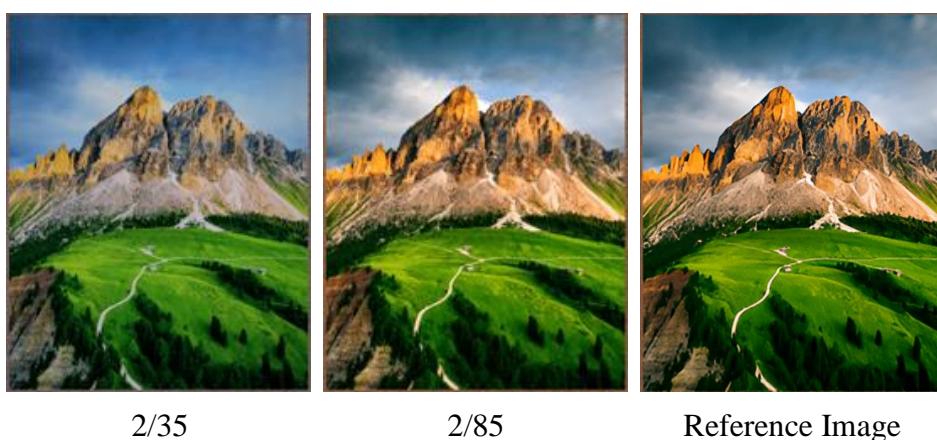
برای ارزیابی عملکرد مدل پیشنهادی بر روی داده‌های جدیدی که تاکنون ندیده است، چهار آزمایش صورت گرفت. در آزمایش نخست عملکرد مدل پیشنهادی بر روی دادگان آزمون از مجموعه داده TID2013 سنجیده شد که در شکل ۱۸-۴ نتیجه ارزیابی مدل پیشنهادی آمده است. مقدار عددی سمت راست هر تصویر مقدار پیش‌بینی مدل پیشنهادی و مقدار سمت چپ مقدار MOS را نشان می‌دهد.

در آزمایش دوم، یک تصویر طبیعی که اعوجاج تاری بر روی آن اعمال شده است و یک تصویر که کنتراست آن تغییر یافته است توسط مدل پیشنهادی ارزیابی گردید. شکل ۱۹-۴ عملکرد مدل

پیشنهادی بر روی دو تصویر اعوجاج یافته را نشان می‌دهد. همان‌طور که انتظار می‌رود تصویر باکیفیت‌تر نمره کیفی بالاتری را توسط مدل کسب کرده است.



شکل ۱۸-۴ : پیش‌بینی مدل پیشنهادی بر روی ۹ تصویر آزمون مجموعه داده TID2013



شکل ۱۹-۴ : عملکرد مدل پیشنهادی بر روی دو تصویر طبیعی اعوجاج یافته

ارزیابی یک مدل بر روی یک مجموعه دادگان دیگر می‌تواند عملکرد آن را به چالش بکشد. مدل پیشنهادی بر روی مجموعه داده TID2013 آموزش دیده است. به دلیل اینکه مجموعه داده KADID10K از نظر گستردگی اعوجاج‌ها به مجموعه داده TID20113 نزدیک است، در آزمایش سوم از مجموعه داده KADID10K استفاده گردید. شکل ۲۲-۴ عملکرد مدل پیشنهادی بر روی تصاویری از مجموعه داده KADID10K که اعوجاج تاری گوسی در پنج شدت متفاوت بر تصویر اصلی متناظر آن‌ها اعمال شده است و شکل ۲۳-۴ پیش‌بینی مدل پیشنهادی برای تصاویری که تغییر کنتراست یافته تصویر مرجع در پنج سطح هستند را نشان می‌دهد.

مقدار DMOS متناظر هر تصویر اعوجاج یافته که برچسب اصلی تصاویر مجموعه داده KADID10K است در زیر هر تصویر آمده است. مدل پیشنهادی بر اساس مقادیر MOS برچسب اصلی تصاویر مجموعه داده TID2013 میزان کیفیت را پیش‌بینی می‌کند. به دلیل متفاوت بودن ماهیت امتیازدهی به روش MOS و DMOS نمی‌توان از مقایسه مقادیر آن‌ها در مورد عملکرد مدل پیشنهادی نتیجه‌ای گرفت. آنچه اهمیت دارد این است که با مقایسه دو تصویر اعوجاج یافته مقدار DMOS و پیش‌بینی مدل برای تصویر باکیفیت پایین‌تر نسبت به تصویری که از کیفیت بیشتری برخوردار است، کمتر باشد؛ به عبارت دیگر با افت کیفیت تصویر هر دو مقادیر DMOS و پیش‌بینی مدل روند نزولی داشته باشند.

به منظور بررسی عملکرد مدل پیشنهادی آزمایش چهارم بر روی تصاویر غیرطبیعی از تصاویر متنی انجام شد. سیستم تشخیص خودکار کیفیت تصویر متنی می‌تواند از ورود تصاویر بدون کیفیت که عملکرد سیستم تشخیص کاراکتر نوری<sup>۱</sup> را پایین می‌آورند جلوگیری کند و یا در سیستم ضبط ویدئو فریم‌های باکیفیت تصویر را انتخاب کند.

شکل ۲۲-۴ دو تصویر متنی از مجموعه داده SoC [۷۳] که از طریق اعمال تاری فوکوس در دو سطح افت کیفیت پیدا کرده‌اند را نشان می‌دهد که مقدار پیش‌بینی مدل پیشنهادی به عنوان کیفیت هر تصویر در زیر هر تصویر مشخص است. شکل ۲۳-۴ عملکرد مدل پیشنهادی بر روی چهار

تصویر اعوجاج یافته دو اعوجاج تاری و نویز گوسی در دو سطح مختلف نشان می‌دهد. مدل پیشنهادی برای تصویر با کیفیت پایین‌تر نسبت به تصویر با کیفیت بیشتر، نمره کمتری پیش‌بینی کرده است.



3/81  
(DMOS = 4/57 )



Reference Image



3/22  
(DMOS = 2/67 )



3/69  
(DMOS = 4/33 )



1/61  
(DMOS = 1/10 )



2/25  
(DMOS = 1/67)

شکل ۲۰-۴ : عملکرد مدل پیشنهادی بر روی تصاویر KADID10K متاثر از اعوجاج تاری گوسی



6/29

(DMOS = 3/33)

Reference Image



5/98

(DMOS = 4/30)

6/15

(DMOS = 3/73)



4/51

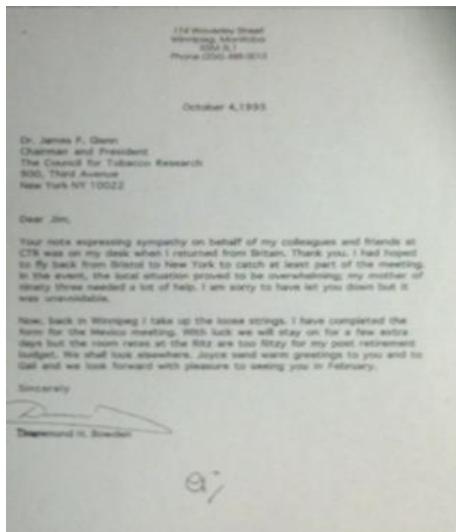
(DMOS = 3/27 )

5/15

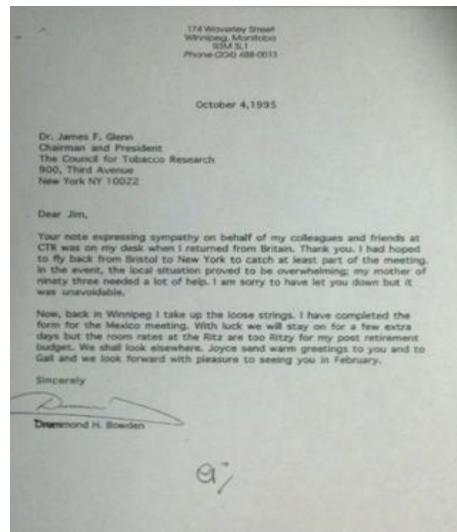
(DMOS = 4/13)

شکل ۲۱-۴ : عملکرد مدل پیشنهادی بر روی تصاویر KADID10K متاثر از تغییر کنترast

## فصل چهارم: نتایج، تجزیه و تحلیل و تفسیر آنها

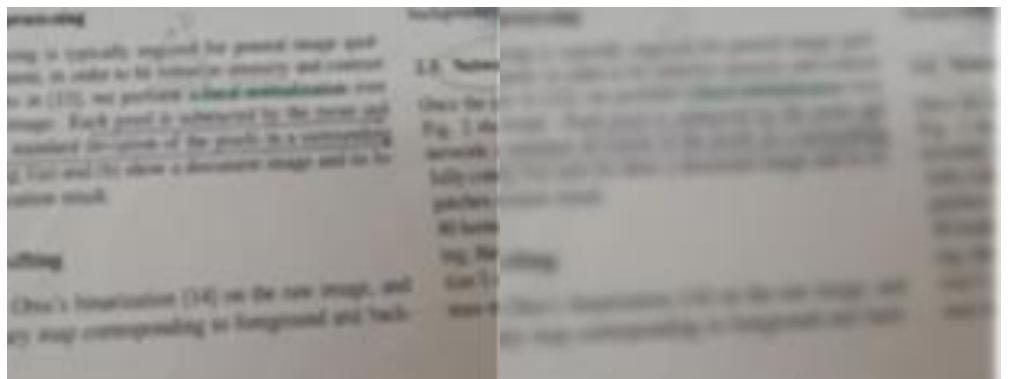


3/27



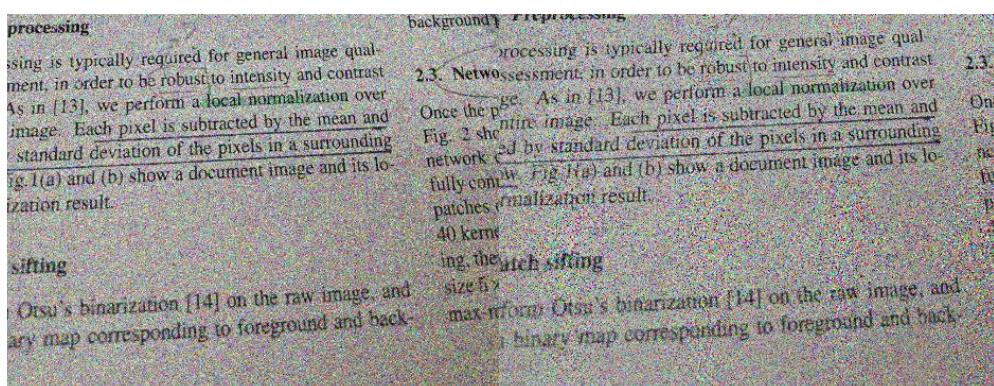
3/46

شکل ۲۲-۴ : عملکرد مدل پیشنهادی بر روی دو تصویر متنی متاثر از تاری فوکوس



2/72

2/02



3/13

3/03

شکل ۲۳-۴ : عملکرد مدل پیشنهادی بر روی تصاویر متنی متاثر از اعوجاج تاری و نویز گوسی

به منظور مقایسه نسخه‌های مختلف شبکه عصبی عمیق ResNet و عملکرد آن‌ها در حوزه ارزیابی کیفیت تصویر، سه معماری ResNet50، ResNet32 و ResNet18 در ساختار مدل پیشنهادی بررسی شدند. عملکرد ResNet50، ResNet32، ResNet18 بر اساس ریشه میانگین مربعات خطأ در جدول ۵-۴ آمده است. استفاده از معماری ResNet18 به عنوان ستون فقرات مدل پیشنهادی نسبت به ResNet50 و ResNet32، از عملکرد بهتری برخوردار است.

جدول ۵-۴ : عملکرد سه نسخه ResNet در ارزیابی کیفیت تصویر

ResNet50	ResNet34	ResNet18	
۰/۶۹	۰/۵۶	۰/۵۴	RMSE

به منظور مقایسه روش پیشنهادی با سایر روش‌های NR-IQA جدول ۶-۴ عملکرد روش پیشنهادی را روش‌های دیگر که بر روی پایگاه داده TID2013 آزمایش شده‌اند، بر اساس دو معیار ارزیابی PLCC و SROCC گزارش می‌کند.

جدول ۶-۴ : مقایسه مدل پیشنهادی با سایر روش‌ها بر روی پایگاه داده TID2013

SROCC	PLCC	محققان	روش
۰/۹۴۶	۰/۹۴۵	مدل راشنانی و همکاران [۱۳]	FR-IQA
۰/۸۲	۰/۷۸	مدل یان و همکاران [۱۴]	
۰/۷۵	-	مدل طالبی و میلانفر [۳۸]	NR-IQA
۰/۹۱۰	۰/۹۱۸	مدل لامیچه و همکاران [۳۷]	
۰/۹۹۸	۰/۹۰	مدل پیشنهادی	

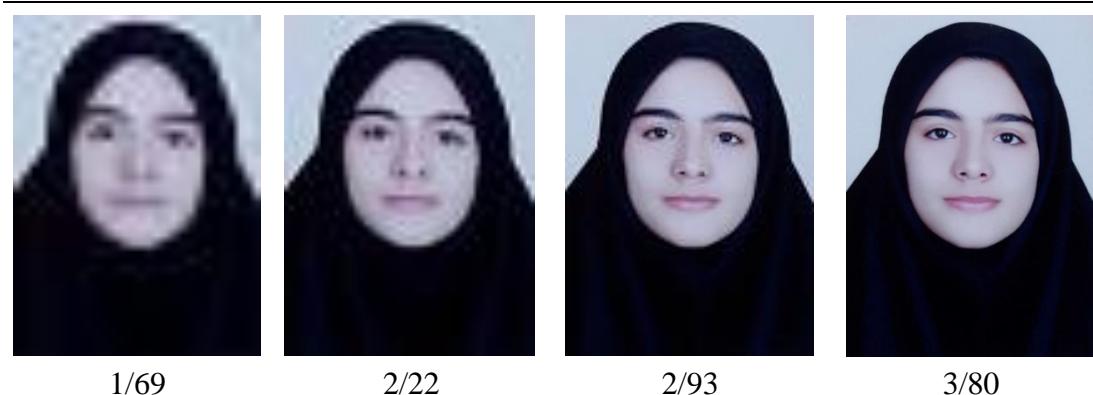
#### ۶-۴ تحلیل عملکرد مدل پیشنهادی در نمونه کاربردی

اکثر سیستم‌های الکترونیکی مثل سامانه سنجش کشوری برای ارائه خدمات، کاربر را ملزم به بارگذاری اطلاعات هویتی نظیر تصاویر کارت ملی و صفحات شناسنامه می‌کنند. این بخش به کاربرد مدل پیشنهادی در چنین سیستم‌هایی می‌پردازد.

فرایند ثبت نام کنکور سراسری در سامانه سنجش را در نظر بگیرید. کاربر می‌بایست در بخش ثبت نام، مشخصات خود را ثبت و تصویر پرسنلی، کارت ملی و تصویر صفحات شناسنامه را در قسمت‌های مشخص شده بارگذاری کند. اگر تصویر پرسنلی قابل شناسایی نباشد و یا تصاویر کارت ملی و صفحات شناسنامه از خوانایی لازم برخوردار نباشند فرایند اعتبارسنجی و تطابق مشخصات ثبت شده با اسناد بارگذاری شده قابل انجام نیست. به کمک مدل پیشنهادی می‌توان میزان خوانایی و کیفیت اسناد بارگذاری شده را بلافاصله پس از بارگذاری، ارزیابی و در صورت نیاز کاربر را ملزم به بارگذاری مجدد اطلاعات نمود.

مدل پیشنهادی میزان کیفیت تصویر را به صورت عددی تعیین می‌کند. به منظور استفاده از آن در سیستم‌های الکترونیکی باید آستانه مناسب به منظور تایید یا عدم تایید یک تصویر از نظر کیفیت تعیین شود. به طور کلی باید حد آستانه را با توجه به اعوجاج خاصی که اغلب هر سیستم با آن رو به رو است تعیین شود. به طور معمول تصاویری که کاربر بارگذاری می‌کند با دوربین تلفن همراه اخذ شده است و ممکن است متأثر از نور زیاد یا تاری حرکتی باشد.

یکی از عوامل افت کیفیت، تاری تصویر ناشی از کاهش حجم آن است. با توجه به تنظیمات هر سامانه تصاویر باید از حجم مناسبی هنگام بارگذاری برخوردار باشند. ممکن است با کم کردن حجم تصاویر به روش‌های مختلف کیفیت آن کاهش یابد و حتی دیگر قابل شناسایی و استفاده نباشد. ارزیابی خودکار کیفیت تصویر در هنگام بارگذاری از پذیرش چنین تصاویری جلوگیری می‌کند. چهار تصویر پرسنلی در شکل ۲۴-۴ آمده است. به ترتیب از راست به چپ حجم تصاویر کاهش یافته است و به تدریج تاری ناشی از کاهش حجم منجر به افت کیفیت شده است. نتیجه ارزیابی کیفیت این تصاویر توسط مدل پیشنهادی نشان می‌دهد که تصویری که مقدار کمی کیفیت آن ۲/۹۳ است قابل شناسایی است ولی تصویر که کیفیت آن ۲/۲۲ است را به سختی می‌توان تطابق داد. در نتیجه می‌توان مقدار ۲/۵ را به عنوان آستانه پذیرش سامانه برای تصاویر پرسنلی در نظر گرفت. به عبارت دیگر تصاویری که دارای کیفیت کمتر از ۲/۵ باشند سامانه باید درخواست بارگذاری تصویر جایگزین را به کاربر بدهد.



شکل ۲۴-۴ : عملکرد مدل پیشنهادی بر روی تصاویر پرسنلی



شکل ۲۵-۴ : عملکرد مدل پیشنهادی بر روی تصاویر کارت ملی

در شکل ۲۵-۴ چند نمونه تصویر کارت ملی خوانا و نسخه‌های اعوجاج یافته آنها که به دلیل

نور زیاد و یا تاری ناشی از حرکت دست خوانا نیستند را نشان می‌دهد. آستانه تشخیص قابل قبول بودن تصویر بارگذاری شده در سیستم را می‌توان از مقایسه پیش‌بینی‌های مدل پیشنهادی برای تصاویر اعوجاج یافته و تصاویر قابل قبول در نظر گرفت. نتایج حاصل از ارزیابی تصاویر اعوجاج یافته توسط مدل پیشنهادی نشان می‌دهد که تصاویر متاثر از تاری حرکتی با اینکه ناخوانا هستند امتیاز کیفی بیشتری را نسبت به تصاویری که متاثر از نور زیاد هستند و به نسبت خواناتر کسب کرده‌اند. می‌توان نتیجه گرفت مدل پیشنهادی نسبت به نور حساس‌تر است و تصاویری که با نورپردازی مناسب اخذ شده‌اند را با کیفیت‌تر تلقی می‌کند. در اینجا اگر بخواهیم سخت‌گیرانه عمل کنیم می‌توان حد آستانه را ۴ در نظر گرفت.

در گام بعدی هدف بررسی امکان استفاده از حد آستانه یکسان برای پذیرش دو نوع تصاویر برای مثال تصاویر کارت ملی و تصاویر شناسنامه است. در شکل ۲۶-۴ ارزیابی کیفیت تصویر صفحه اول شناسنامه خوانا و تصویری که متاثر از تاری حرکتی است را نشان می‌دهد. نرخ ارزیابی تصاویر در شکل ۲۶-۴ و شکل ۲۵-۴ یکسان نیست. می‌توان نتیجه گرفت تشخیص مدل پیشنهادی به نوع تصویر بستگی دارد و تعیین آستانه مجزا لازم است.



5/70



4/23

شکل ۲۶-۴ : عملکرد مدل پیشنهادی بر روی تصاویر شناسنامه

در کاربرد عملی قسمت بارگذاری هر نوع سند مجزا است و می‌توان حد آستانه مدل پیشنهادی

را به صورت مجزا تنظیم نمود. البته در صورتی که بخواهیم اسناد به صورت یکپارچه ارزیابی شوند می‌توان با استفاده از یک مدل طبقه‌بندی تصویر کارت ملی و شناسنامه را تشخیص داد سپس هر نوع تصویر را توسط مدل پیشنهادی با حد آستانه ویژه ارزیابی نمود.

## **فصل ۵:**

**نتیجه‌گیری و پیشنهادها**

در این فصل روش پیشنهادی برای ارزیابی خودکار کیفیت تصویر بدون مرجع به صورت خلاصه توضیح داده می‌شود. نتایج حاصل از این پژوهش با سایر روش‌ها مقایسه می‌شوند و پیشنهادهایی به منظور پژوهش‌های آینده در حوزه تحقیق بیان می‌شود.

### ۱-۵ جمع‌بندی

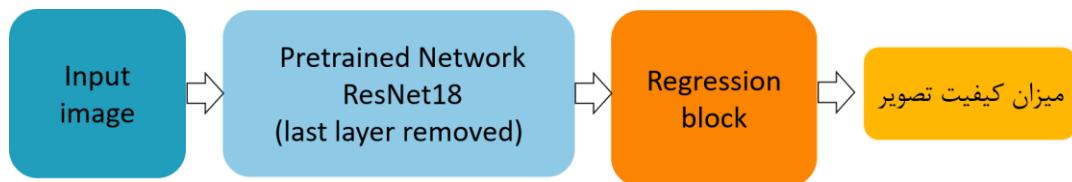
پژوهش‌های پیشین در زمینه ارزیابی کیفیت تصویر بررسی شدند. مدل‌های موفقی در پیش‌بینی کیفیت تصویر نزدیک به نظر انسان ارائه شده بود. با توجه به اهمیت و کارایی تصاویر و لزوم ارزیابی کیفیت تصویر بدون نیاز به مرجع آن نیاز به بهبود راه حل‌های موجود احساس شد. در این پژوهش یک مدل ارزیابی خودکار کیفیت تصویر ارائه گردید. شبکه از پیش آموزش دیده ResNet18 به عنوان ستون فقرات مدل پیشنهادی استفاده گردید. بلوك رگرسیون برای ارزیابی کیفیت تصویر طراحی شد. پایگاه داده TID2013 برای آموزش مدل به کار گرفته شد. شکل ۱-۵ مدل پیشنهادی مبتنی بر یادگیری انتقالی و بازنظمی شبکه از پیش آموزش دیده ResNet18 برای به کار گیری در مسئله ارزیابی خودکار کیفیت تصویر بدون مرجع را نشان می‌دهد.

عملکرد مدل پیشنهادی بر روی تصاویری از مجموعه دادگان TID2013، KADID10K و تصاویر متنی ارزیابی شد. همچنین در یک سیستم الکترونیکی عملکرد مدل پیشنهادی در مورد کاربردی پذیرش کارت ملی بررسی شد. نتایج به دست آمده در رویکرد مورد نظر پایان‌نامه رضایت‌بخش بود. سایر مدل‌های ارائه شده در حوزه ارزیابی کیفیت تصویر با مدل پیشنهادی بر اساس سه معیار ارزیابی PLCC، SROCC و RMSE مقایسه شدند.

### ۲-۵ نتیجه‌گیری

نتایج تجربی اثربخشی معماری شبکه پیشنهادی را تأیید می‌کند و نشان می‌دهد که مدل پیشنهادی عملکرد مطلوبی دارد و نسخه تحریف‌شده تصاویر را به درستی ارزیابی می‌کند. آزمایش‌های انجام‌شده نشان داد که مدل پیشنهادی در مقایسه با روش‌های دیگر ارائه شده در این حوزه عملکرد قابل قبولی دارد. سه نسخه مختلف شبکه ResNet به عنوان ستون فقرات مدل پیشنهادی بررسی شدند. نتایج نشان داد که برای مسئله این پژوهش استفاده از یک شبکه با ساختار ساده‌تر و پارامترهای کمتر

ResNet18 نسبت به ResNet32 و ResNet52 که پیچیده‌تر هستند، عملکرد بهتری را نتیجه می‌دهد. مشاهده شد که روش‌های یادگیری انتقالی می‌تواند به عنوان یک روش کاربردی و مطمئن برای ارزیابی کیفیت یک تصویر به طور خودکار و بدون تصویر مرجع استفاده گردد.



شکل ۱-۵ : مدل پیشنهادی

### ۳-۵ پیشنهادها

در این پژوهش از نسخه‌های مختلف شبکه ResNet برای آزمایش‌ها استفاده گردید. برای پژوهش‌های آینده پیشنهاد می‌شود سایر شبکه‌های از پیش آموزش داده شده به عنوان ستون فقرات مدل پیشنهادی بررسی شوند. ساختار بلوك رگرسیون نیز می‌تواند تغییر داده شود و به صورت پیچیده‌تر طراحی شود. مدل پیشنهادی بر روی مجموعه داده TID2013 آموزش یافته است. این روش برای مجموعه دادگان مختلف قابل تعمیم است و می‌توان در پژوهش‌های بعدی از سایر مجموعه دادگان برای آموزش و ارزیابی مدل استفاده کرد و قابلیت شبکه را سنجید. استفاده از مجموعه داده KADID10K برای آموزش یک مدل به منابع محاسباتی و پردازشی نیاز دارد که در این پژوهش آموزش مدل پیشنهادی بر روی این مجموعه داده به علت کمبود منابع امکان‌پذیر نبود. در اقدامی دیگر بررسی عملکرد مدل پیشنهادی در ارزیابی کیفی اسناد متنوع نظری تصویر گواهی‌نامه در نمونه‌های کاربردی متنوع می‌تواند در پژوهش‌های آتی مدنظر قرار گیرد.

## مراجع و مأخذ

- [1] A. Punchihewa and D. G. Bailey, "Artefacts in image and video systems; classification and mitigation," in *Proceedings of image and vision computing New Zealand*, 2002, pp. 197-202.
- [2] J. Kim and S. Lee, "Deep learning of human visual sensitivity in image quality assessment framework," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1676-1684.
- [3] J. Kim and S. Lee, "Fully deep blind image quality predictor," *IEEE Journal of selected topics in signal processing*, vol. 11, no. 1, pp. 206-220, 2016.
- [4] J. Kim, A.-D. Nguyen, and S. Lee, "Deep CNN-based blind image quality predictor," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 1, pp. 11-24, 2018.
- [5] D. Liang, X. Gao, W. Lu, and J. Li, "Deep blind image quality assessment based on multiple instance regression," *Neurocomputing*, vol. 431, pp. 78-89, 2021.
- [6] P. R. Rajarapollu and V. R. Mankar, "Bicubic interpolation algorithm implementation for image appearance enhancement," *Ij cst*, vol. 8, no. 2, pp. 23-26, 2017.
- [7] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE transactions on Image Processing*, vol. 21, no. 8, pp. 3339-3352, 2012.
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600-612, 2004.
- [9] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386, 2011.
- [10] S. A. Amirshahi, M. Pedersen, and S. X. Yu, "Image quality assessment by comparing CNN features between images," *Journal of Imaging Science and Technology*, vol. 60, no. 6, pp. 60410-1-60410-10, 2016.
- [11] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1733-1740.
- [12] Y. Li, X. Ye, and Y. Li, "Image quality assessment using deep convolutional networks," *AIP Advances*, vol. 7, no. 12, p. 125324, 2017.
- [13] E. Prashnani, H. Cai, Y. Mostofi, and P. Sen, "Pieapp: Perceptual image-error assessment through pairwise preference," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1808-1817.
- [14] Q. Yan, D. Gong, and Y. Zhang, "Two-stream convolutional networks for blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2200-2211, 2018.
- [15] T. Jiang , X. Hu, X. Yao, L. Tu, J. Huang, X. Ma, J. Cui, Q. Wu, and J. Xu, "Tongue image quality assessment based on a deep convolutional neural network," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1-14, 2021.
- [16] J. Su , M. Li, Y. Lin, L. Xiong, C. Yuan, Z. Zhou, and K. Yan, "A deep learning-based automatic image quality assessment method for respiratory phase on computed tomography chest images," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 3, p. 2240, 2024.

- [17] N. Mamat, M. F. Othman, R. Abdoulghafar, S. B. Belhaouari, N. Mamat, and S. F. Mohd Hussein, "Advanced technology in agriculture industry by implementing image annotation technique and deep learning approach: A review," *Agriculture*, vol. 12, no. 7, p. 1033, 2022.
- [18] S. M. Khorashadizadeh, M. SadeghiBakhi, F. Seifishahpar, and A. Latif, "Blind image quality assessment based on Multi-resolution Local Structures," in *11th International Conference on Computer Engineering and Knowledge (ICCKE)*, 2021: IEEE, pp. 310-315.
- [19] M. Behzadpour and M. Ghanbari, "Improving precision of objective image/video quality meters," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 4465-4478, 2023.
- [20] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98-117, 2009.
- [21] B. Girod, "What's wrong with mean-squared error?," *Digital images and human vision*, pp. 207-220, 1993.
- [22] R. A. Manap and L. Shao, "Non-distortion-specific no-reference image quality assessment: A survey," *Information Sciences*, vol. 301, pp. 141-160, 2015.
- [23] A. Rehman and Z. Wang, "Reduced-reference image quality assessment by structural similarity estimation," *IEEE transactions on image processing*, vol. 21, no. 8, pp. 3378-3389, 2012.
- [24] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE journal of selected topics in signal processing*, vol. 3, no. 2, pp. 202-211, 2009.
- [25] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *Human vision and electronic imaging X*, 2005, vol. 5666: SPIE, pp. 149-159.
- [26] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized DCT-based image representation," *IEEE Transactions on multimedia*, vol. 13, no. 4, pp. 824-829, 2011.
- [27] Y. Liang, J. Wang, X. Wan, Y. Gong, and N. Zheng, "Image quality assessment using similar scene as reference," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14*, 2016: Springer, pp. 3-18.
- [28] K. Chono, Y.-C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality assessment using distributed source coding," in *2008 IEEE International Conference on Multimedia and Expo*, 2008: IEEE, pp. 609-612.
- [29] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, and W. Gao, "Reduced-reference image quality assessment in free-energy principle and sparse representation," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 379-391, 2017.
- [30] R. Soundararajan and A. C. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517-526, 2011.
- [31] K. Gu, G. Zhai, X. Yang, and W. Zhang, "A new reduced-reference image quality assessment using structural degradation model," in *2013 IEEE international symposium on circuits and systems (ISCAS)*, 2013: IEEE, pp. 1095-1098.

- [32] X. Li, "Blind image quality assessment," in *Proceedings. International Conference on Image Processing*, 2002, vol. 1: IEEE, pp. I-I.
- [33] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs. non-handcrafted features for computer vision classification," *Pattern Recognition*, vol. 71, pp. 158-172, 2017.
- [34] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850-4862, 2014.
- [35] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE transactions on Image Processing*, vol. 20, no. 12, pp. 3350-3364, 2011.
- [36] P. Bovik, R. Soundararajan, and C. Bampis, "Laboratory for image & video engineering," ed, 2009.
- [37] K. Lamichhane, M. Carli, and F. Battisti, "A CNN-based no reference image quality metric exploiting content saliency," *Signal Processing: Image Communication*, vol. 111, pp. 116-899, 2023.
- [38] H. Talebi and P. Milanfar, "NIMA: Neural image assessment," *IEEE transactions on image processing*, vol. 27, no. 8, pp. 3998-4011, 2018.
- [39] A. G. Howard, "Mobileneets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.
- [42] X. Peng, H. Cao, and P. Natarajan, "Document image OCR accuracy prediction via latent Dirichlet allocation," in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*, 2015: IEEE, pp. 771-775.
- [43] X. Peng, H. Cao, and P. Natarajan, "Document image quality assessment using discriminative sparse representation," in *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*, 2016: IEEE, pp. 227-232.
- [44] T. Obafemi-Ajayi and G. Agam, "Character-based automated human perception quality assessment in document images," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 42, no. 3, pp. 584-595, 2011.
- [45] W. Hou, X. Gao, D. Tao, and X. Li, "Blind image quality assessment via deep learning," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 6, pp. 1275-1286, 2014.
- [46] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," *Signal, Image and Video Processing*, vol. 12, pp. 355-362, 2018.
- [47] H. Li, J. Qiu, and F. Zhu, "TextNet for Text-Related Image Quality Assessment," in *Artificial Neural Networks and Machine Learning–ICANN 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, Proceedings, Part II* 27, 2018: Springer, pp. 275-285.
- [48] A. Alaei, D. Conte, M. Martineau, and R. Raveaux, "Blind document image quality prediction based on modification of quality aware clustering method integrating a patch selection strategy," *Expert Systems with Applications*, vol. 108, pp. 183-192, 2018.

- [49] H. Li, F. Zhu, and J. Qiu, "CG-DIQA: No-reference document image quality assessment based on character gradient," in *24th International Conference on Pattern Recognition (ICPR)*, 2018: IEEE, pp. 3622-3626.
- [50] L. Kang, P. Ye, Y. Li, and D. Doermann, "A deep learning approach to document image quality assessment," in *IEEE International Conference on Image Processing (ICIP)*, 2014: IEEE, pp. 2570-2574.
- [51] P. Li, L. Peng, J. Cai, X. Ding, and S. Ge, "Attention based RNN model for document image quality assessment," in *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, 2017, vol. 1: IEEE, pp. 819-825.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
- [53] X. Ou, P. Yan, Y. Zhang, B. Tu, G. Zhang, J. Wu, and W. Li., "Moving object detection method via ResNet-18 with encoder-decoder structure in complex scenes," *IEEE Access*, vol. 7, pp. 108152-108160, 2019.
- [54] S. Ioffe, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [55] L. N. Smith, "No more pesky learning rate guessing games," *CoRR, abs/1506.01186*, vol. 5, p. 575, 2015.
- [56] J. Howard and S. Gugger, "Fastai: a layered API for deep learning," *Information*, vol. 11, no. 2, p. 108, 2020.
- [57] D. Chandler and S. Hemami, "A57 database," ed, 2007.
- [58] A. Roman-Gonzalez, "The SNCD as a Metrics for Image Quality Assessment," *International journal of advanced computer science and applications (IJACSA)*, vol. 4, no. 7, pp. 40-47, 2013.
- [59] H. Sheikh, "LIVE image quality assessment database release 2," <http://live.ece.utexas.edu/research/quality>, 2005.
- [60] Z. Zhang, H. Wang, S. Liu, and T. S. Durrani, "Deep activation pooling for blind image quality assessment," *Applied Sciences*, vol. 8, no. 4, p. 478, 2018.
- [61] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of electronic imaging*, vol. 19, no. 1, pp. 011006-011006-21, 2010.
- [62] J. J. M. Escobar, O. M. Matamoros, I. L. Reyes, R. T. Padilla, and L. C. Hernández, "Defining a no-reference image quality assessment by means of the self-affine analysis," *Multimedia Tools and Applications*, vol. 80, no. 9, pp. 14305-14320, 2021.
- [63] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of modern radioelectronics*, vol. 10, no. 4, pp. 30-45, 2009.
- [64] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.C. Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal processing: Image communication*, vol. 30, pp. 57-77, 2015.
- [65] K. Gu, G. Zhai, X. Yang, W. Zhang, and M. Liu, "Subjective and objective quality assessment for images with contrast change," in *2013 IEEE International Conference on Image Processing*, 2013: IEEE, pp. 383-387.

- [66] K. Gu, G. Zhai, W. Lin, and M. Liu, "The analysis of image contrast: From quality assessment to automatic enhancement," *IEEE transactions on cybernetics*, vol. 46, no. 1, pp. 284-297, 2015.
- [67] T. Virtanen, M. Nuutinen, J. Radun, T. M. Leisti, and J. P. Häkkinen, "Alternative performance metrics and target values for the CID2013 database," in *Image Quality and System Performance XII*, 2015, vol. 9396: SPIE, pp. 233-243.
- [68] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A large-scale artificially distorted IQA database," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019: IEEE, pp. 1-3.
- [69] M. Frackiewicz, G. Szolc, and H. Palus, "An improved SPSIM index for image quality assessment," *Symmetry*, vol. 13, no. 3, p. 518, 2021.
- [70] N. Murray, L. Marchesotti, and F. Perronnin, "AVA: A large-scale database for aesthetic visual analysis," in *2012 IEEE conference on computer vision and pattern recognition*, 2012: IEEE, pp. 2408-2415.
- [71] J. Ruikar and S. Chaudhury, "NITS-IQA Database: A New Image Quality Assessment Database," *Sensors*, vol. 23, no. 4, p. 2279, 2023.
- [72] J. Antkowiak, T.J. Baina, F.V. Baroncini, N. Chateau, F. FranceTelecom, A.C. Pessoa, F.S. Colonnese, I.L. Contin, J. Caviedes, and F. Philips "Final report from the video quality experts group on the validation of objective models of video quality assessment march 2000," *Final report from the video quality experts group on the validation of objective models of video quality assessment march*, vol. 10, 2000.
- [73] J. Kumar, P. Ye, and D. Doermann, "A dataset for quality assessment of camera captured document images," in *Camera-Based Document Analysis and Recognition: 5th International Workshop, CBDAR 2013, Washington, DC, USA, August 23, 2013, Revised Selected Papers 5*, 2014: Springer, pp. 113-125.
- [74] "<Drone-based object counting by spatially regularized regional proposal network.pdf>."
- [75] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
- [76] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84-90, 2017.
- [77] L. Cheng, A. C. S. Leung, and S. Ozawa, *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part III*. Springer, 2018.
- [78] L. Zhu, Z. Li, C. Li, J. Wu, and J. Yue, "High performance vegetable classification from images based on alexnet deep learning model," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 4, pp. 217-223, 2018.
- [79] A. Alsayed, A. Alsabei, and M. Arif, "Classification of apple tree leaves diseases using deep learning methods," *International Journal of Computer Science & Network Security*, vol. 21, no. 7, pp. 324-330, 2021.
- [80] X. Meng, Y. Yuan, G. Teng, and T. Liu, "Deep learning for fine-grained classification of jujube fruit in the natural environment," *Journal of Food Measurement and Characterization*, vol. 15, no. 5, pp. 4150-4165, 2021.

Λ+

## پیوست‌ها

### یادگیری ماشین

یادگیری ماشین زیرمجموعه‌ای از هوش مصنوعی است که به سیستم توانایی یادگیری خودکار و بهبود تجربه بدون برنامه‌ریزی صریح را می‌دهد. یادگیری ماشین بر توسعه برنامه‌های رایانه‌ای متمرکز است که می‌تواند به داده‌ها دسترسی داشته و از آن‌ها برای یادگیری خود استفاده کنند. فرآیند یادگیری برای یک ماشین با مشاهده‌ها یا داده‌ها مانند مثال‌ها، تجربه مستقیم یا دستورالعمل‌ها آغاز می‌شود و بر اساس الگوهایی که به ماشین داده می‌شود، به دنبال الگوهای نهفته در داده‌ها می‌گردد. هدف اصلی یادگیری ماشین این است که به کامپیوترها اجازه داده شود بدون دخالت یا کمک انسان به طور خودکار یاد بگیرند و مجموعه اقدام‌ها را بر اساس آن تنظیم کنند. اگر سیستمی با دریافت اطلاعات و داده از محیط اطراف توانایی بهبود عملکرد خود را داشته باشد گفته می‌شود یادگیری توسط آن سیستم صورت گرفته و سیستم به صورت خودکار توانایی حل مسئله و تصمیم‌گیری را دارد.

روش‌های یادگیری ماشین در دو مرحله آموزش و آزمایش انجام می‌گیرد. در مرحله آموزش یک مجموعه داده آموزشی به همراه خروجی مطلوب در اختیار الگوریتم یادگیری ماشین قرار می‌گیرد تا یک مدل بر پایه آن تولید شود. در مرحله آزمایش یک مجموعه داده‌های آزمایشی که از داده‌های مرحله آموزش مجزا هستند در اختیار مدل تولیدشده قرار می‌گیرد تا میزان کارایی آن از طریق قیاس خروجی مدل و خروجی مطلوب متناظر دادگان آزمایشی سنجیده شود.

یادگیری ماشین تجزیه و تحلیل مقادیر عظیم داده را امکان‌پذیر می‌کند و کمک فراوانی به صرفه‌جویی در هزینه‌های عملیاتی و بهبود سرعت عمل در تجزیه و تحلیل داده‌ها می‌کند. تلفیق یادگیری ماشین و تکنولوژی‌های پیشرفته امروز باعث می‌شود که اهمیت یادگیری ماشین در همه ابعاد زندگی انسان دیده شود. از کاربردهای یادگیری ماشین می‌توان به شناسایی تصویر، تحلیل احساسات و تحلیل گفتار اشاره کرد.

در ادامه مختصری از چهار روش یادگیری در این حوزه یادگیری ماشین که شامل یادگیری

با ناظر، یادگیری بدون ناظر، یادگیری نیمه نظارتی و یادگیری تقویتی مطرح می‌شود.

### یادگیری با ناظر

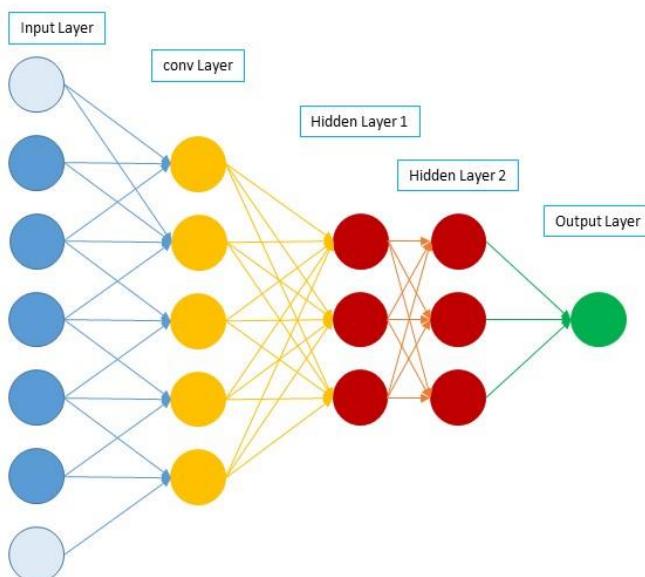
الگوریتم‌های یادگیری ماشین با ناظر برای یادگیری از روی نمونه‌ها طراحی شده‌اند. داده‌های آموزشی هنگام آموزش الگوریتم یادگیری با ناظر مت Shank از ورودی‌هایی است که با خروجی‌های صحیح جفت شده است؛ یعنی داده‌ها دارای برچسب خروجی‌های صحیح هستند. در حین آموزش، الگوریتم الگوهایی در داده‌ها را جست‌وجو می‌کند که با خروجی‌های مدنظر ارتباط دارند و آن الگوها را یاد می‌گیرد. الگوریتم یادگیری می‌تواند خروجی خود را با خروجی صحیح مقایسه کرده و میزان خطای محاسبه کند تا بتواند بر این اساس مدل را اصلاح کند. پس از آموزش، الگوریتم یادگیری با ناظر ورودی‌های جدیدی را می‌گیرد که خروجی مدنظر آن‌ها مشخص نیست. بر اساس داده‌های آموزشی قبلی تعیین می‌کند که ورودی‌های جدید به کدام برچسب طبقه‌بندی شوند یا به عبارت دیگر چه خروجی دارند. هدف از یک مدل یادگیری با ناظر پیش‌بینی برچسب صحیح برای داده‌های ورودی جدید است.

الگوریتم‌های یادگیری با ناظر به دو دسته طبقه‌بندی و رگرسیون تقسیم می‌شوند که از اهمیت زیادی برخوردار هستند. در مسائل طبقه‌بندی، در طول آموزش، داده‌هایی به یک الگوریتم طبقه‌بندی وارد می‌شوند که هر یک از قبل در یک دسته یا کلاس خاصی طبقه‌بندی شده است. وظیفه‌ی یک الگوریتم طبقه‌بندی این است که بعد از فرایند آموزش یک مقدار ورودی بگیرد و آن را به یک کلاس یا دسته اختصاص دهد که بر اساس چیزی که در حین آموزش یاد گرفته است، آن داده در آن کلاس قرار بگیرد. در الگوریتم‌های طبقه‌بندی سعی می‌شود تابع نگاشت  $f$  را از متغیرهای ورودی  $x$  به متغیرهای خروجی گستته یا طبقه‌ای  $y$  محاسبه شود. برای مثال، الگوریتمی که تشخیص می‌دهد یک ایمیل اسپم است یا نه، یک مسئله طبقه‌بندی را حل می‌کند.

در مسائل رگرسیون الگوریتم سعی در تخمین مقداری پیوسته بر اساس داده‌های ورودی دارد. رگرسیون یک فرایند آماری پیش‌بینی‌کننده است که در آن مدل سعی دارد رابطه مهم میان متغیرهای وابسته و مستقل را پیدا و کمیت‌ها یا اندازه‌ها را پیش‌بینی کند. در الگوریتم‌های رگرسیون

سعی در محاسبه تابع نگاشت  $f$  از متغیرهای ورودی  $x$  به متغیرهای عددی یا پیوسته خروجی  $y$  دارند. در این حالت،  $y$  یک مقدار حقیقی است که می‌تواند یک عدد صحیح یا یک نقطه شناور باشد. برای مثال، الگوریتمی که مبتلایان به ویروس کرونا در ماه آینده را تخمین می‌زند، یک نمونه رگرسیون است. رگرسیون خطی، ساده‌ترین و شناخته‌شده‌ترین مدل رگرسیون برای تخمین یک مقدار پیوسته است.

در شکل ۱ تصویری از معماری شبکه‌های عصبی کانولوشنال برای رگرسیون، نشان داده شده است. لایه آخر که خروجی شبکه را نشان می‌دهد برای مسئله رگرسیون شامل تنها یک گره است زیرا یک مقدار عددی را نتیجه می‌دهد.



شکل ۱: معماری شبکه کانولوشنال برای مسئله رگرسیون

### یادگیری بدون ناظر

این روش یادگیری ماشین شامل الگوریتم‌هایی است که روی داده‌های بدون برچسب آموزش می‌بینند. الگوریتم از طریق مجموعه داده‌هایی که دریافت می‌کند، به دنبال هرگونه ارتباط معنی‌داری بین داده‌ها است. در این نوع از الگوریتم‌ها اطلاعات مورد استفاده برای آموزش نه طبقه‌بندی شده و نه دارای برچسب هستند. این داده‌ها را کاوش می‌کند و می‌تواند از مجموعه داده‌ها استنباط کند تا ساختارهای پنهان داده‌های بدون برچسب را توصیف کند و ارتباط و الگوی بین داده‌ها را پیدا کند.

دو روش خوشبندی و کاهش بعد در یادگیری بدون ناظر شناخته شده است. در الگوریتم‌های خوشبندی، داده‌ها بر اساس ویژگی‌هایشان، تقسیم‌بندی می‌شوند و داده‌هایی که ویژگی‌های مشابه داشته باشند در یک گروه قرار خواهند گرفت. روش کاهش بعد به کاهش ویژگی داده‌ها می‌پردازد. در یادگیری ماشین یک داده ممکن است چندین ویژگی (چندین بعد) داشته باشد و برخی از ویژگی‌ها تاثیری در یادگیری مدل ندارند برای ساده کردن مسئله می‌توان این ویژگی‌ها را با استفاده از الگوریتم‌های کاهش بعد، حذف کرد. کاهش بعد فضای موردنیاز برای ذخیره داده‌ها را کم خواهد کرد و باعث افزایش سرعت الگوریتم‌های یادگیری خواهد شد.

### یادگیری نیمه نظارتی

یادگیری نیمه نظارت شده میان یادگیری نظارت شده و بدون نظارت قرار می‌گیرد. یادگیری نیمه نظارتی از داده‌های برچسب دار و غیر برچسب برای آموزش استفاده می‌کند. تعداد کمی از دادگان دارای برچسب و تعداد زیادی از آن‌ها بدون برچسب هستند.

### یادگیری تقویتی

یادگیری تقویتی یک روش یادگیری است که از طریق اقدام‌هایی، با محیط خود ارتباط برقرار می‌کند و خطاهای پاداش‌ها را کشف می‌کند و بر اساس این خطاهای پاداش‌ها به هدف خود می‌رسد. جستجوی آزمایش خطا و تأخیر در پاداش، مهم‌ترین ویژگی یادگیری تقویتی است. این روش به سیستم‌ها و عوامل نرم‌افزاری اجازه می‌دهد تا رفتار ایده‌آل را در یک زمینه خاص به‌طور خودکار تعیین کنند تا عملکرد آن به حداقل برسد. بازخورد پاداش ساده برای عامل لازم است تا یاد بگیرد کدام عمل بهتر است و این به عنوان سیگنال تقویت‌کننده شناخته می‌شود.

یکی از مهم‌ترین و کاربردی‌ترین الگوریتم‌ها در یادگیری ماشین به نام الگوریتم گرادیان کاهشی است. این الگوریتم، یک الگوریتم بهینه‌سازی تکراری مرتبه اول است که مینیمم محلی در یک تابع مشتق‌پذیر را پیدا می‌کند. زمانی که پیچیدگی تابعی زیاد باشد معادله مشتق تابع برابر صفر قابل حل نیست و نمی‌توان اکسترمم محلی را یافت. برای حل این مشکل از الگوریتم گرادیان کاهشی استفاده می‌شود. الگوریتم گرادیان کاهشی به این صورت عمل می‌کند؛ در ابتدا یک نقطه به صورت

تصادفی انتخاب می‌شود. سپس مقدار مشتق تابع در آن نقطه محاسبه می‌شود. در مرحله بعد برای رسیدن به نقطه مینیمم محلی، الگوریتم در جهت خلاف مشتق حرکت می‌کند از همین روی است که گرادیان کاهشی به حرکت در خلاف جهت گرادیان اشاره دارد. مطابق رابطه (۱-۲) فرایند تکرار می‌شود تا مینیمم محلی مشخص گردد. در رابطه (۱)  $\eta$  طول گام برای مشتق و  $\frac{df}{dx}$  مشتق تابع  $f$  را نشان میدهد.

$$X_{\text{new}} = X_{\text{old}} - \eta \frac{df}{dx} (X_{\text{old}}) \quad (1)$$

به این دلیل که ممکن است در روند مشتق‌گیری و انتخاب نقطه کاندید جدید نوسان‌هایی صورت گیرد، از ضریبی به عنوان طول گام برای مشتق استفاده می‌شود. این پارامتر که بین ۰ و ۱ است، می‌تواند از قدرت مشتق بکاهد و جلوی حرکت نامنظم را بگیرد. مقدار طول گام که در یادگیری ماشین به عنوان نرخ یادگیری شناخته می‌شود، باید به درستی انتخاب شود تا نتیجه مطلوب حاصل گردد.

یکی از چالش‌های مدل‌های یادگیری ماشین بیش برآش زمانی رخ می‌دهد که یک مدل آموزش‌دهید به صورت دقیق به داده‌های آموزشی خود منطبق می‌شود و سعی می‌کند تمام جزئیات داده‌های آموزشی را حفظ کند به صورتی که با جزئیات و نویزهای موجود در داده‌های آموزشی بیش از حد سازگار می‌شود. به عبارت دیگر مدل در فرآیند یادگیری به اندازه کافی خوب عمومی‌سازی نمی‌کند و در نتیجه نمی‌تواند به طور کلی الگوها و قوانین عمومی را درک کند. وقتی یک مدل دچار بیش برآش می‌شود، عملکرد آن بر روی داده‌های آموزشی به خوبی است؛ اما در مواجهه با داده‌های جدید عملکرد ضعیفی خواهد داشت. استفاده از مدل‌های پیچیده و قدرتمند مانند شبکه‌های عصبی عمیق، می‌تواند به بروز بیش برآش منجر شود. این مدل‌ها توانایی بیشتری برای یادگیری جزئیات و نویزها دارند و ممکن است به جای یادگیری الگوهای عمومی، با جزئیات نویزی در داده‌های آموزشی هماهنگ شوند. به منظور پیش‌گیری از بیش برآش مدل می‌توان اقدام‌های ذیل را اتخاذ نمود.

---

<sup>1</sup>overfitting

## • افزایش تعداد داده‌های آموزشی

می‌توان از طریق تکنیک داده‌افزایی<sup>۱</sup> داده‌های موجود را افزایش داد. مدل با افزایش داده‌ها، بهتر قوانین عمومی را یاد می‌گیرد و توانایی کلیت بخشی بیشتری پیدا می‌کند و به جزئیات نویزی در داده‌ها وابسته نمی‌شود.

## • استفاده از روش‌های منظم‌سازی

روش‌های منظم‌سازی در یادگیری ماشین، روش‌هایی هستند که به منظور کنترل پیچیدگی مدل، جلوگیری از بیش برازش و ایجاد تعمیم‌پذیری بهتر و به طور کلی بهبود عملکرد مدل در داده‌های جدید استفاده می‌شوند. این روش‌ها با کنترل پارامترها پیچیدگی ساختاری مدل را کاهش می‌دهند و به مدل اجازه می‌دهند اطلاعات عمومی را به شکل بهتری درک کند.

یکی از روش‌های منظم‌سازی معروف، استفاده از جریمه کردن است. در این روش، یک جریمه به تابع هدف اضافه می‌شود که بر اساس پارامترهای مدل و وزن‌ها محاسبه می‌شود.

دو نوع جریمه رایج وجود دارد. جریمه مقدار مطلق<sup>۲</sup> و جریمه مربعات.<sup>۳</sup> در جریمه مقدار مطلق، مقدار مجموع مطلق وزن‌ها به تابع هدف اضافه می‌شود. این روش باعث کاهش وزن‌های غیرضروری می‌شود و ویژگی‌های مهم را برجسته می‌کند. در روش جریمه مربعات، مقدار مجموع مربعات وزن‌ها به تابع هدف اضافه می‌شود. این رویکرد باعث تفاوت بین وزن‌ها می‌شود که به پایداری مدل کمک می‌کند.

یک روش دیگر حذف تصادفی<sup>۴</sup> است. در این روش، در هر مرحله از آموزش، تصادفی برخی از واحدهای شبکه عصبی غیرفعال می‌شوند. این تکنیک باعث کاهش اتصال‌های میان واحدها می‌شود و برخی از ویژگی‌های افراطی را نادیده می‌گیرد.

## • استفاده از یک مجموعه اعتبارسنجی متقابل<sup>۵</sup>

به منظور پیشگیری از بیش برازش مدل می‌توان مرحله ارزیابی مدل را در حین آموزش انجام

<sup>1</sup>dataAugmentation

<sup>2</sup>L1 Regularization

<sup>3</sup>L2 Regularization

<sup>4</sup>Dropout

<sup>5</sup>Cross-validation

داد به صورتی که قسمتی از دادگان آموزشی موجود را به عنوان داده اعتبارسنج در نظر گرفت. به این ترتیب که مدل بر روی بخشی از داده‌ها آموزش می‌بیند و سپس بر روی بخش دیگری اعتبارسنجی می‌شود.

شرح حوزه‌های دیگر هوش مصنوعی در ادامه آمده است.

### بینایی کامپیووتر

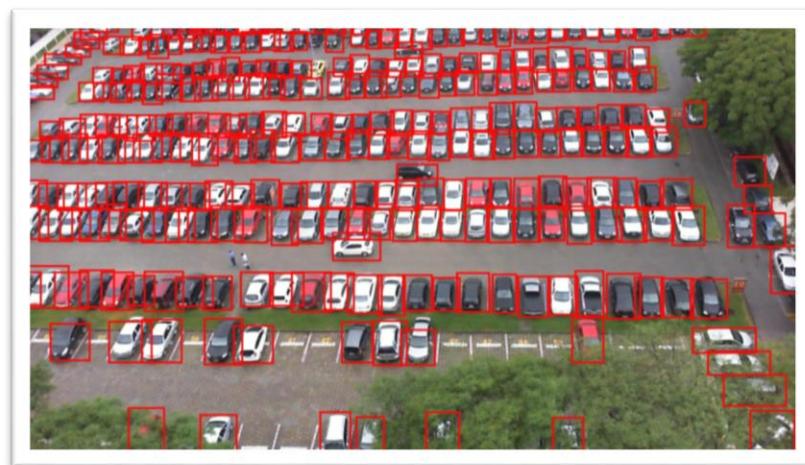
بینایی کامپیووتر حوزه‌ای از هوش مصنوعی مبتنی بر پردازش تصاویر است. امروزه سیستم‌های بینایی کامپیووتر از طریق اصلاح و بهینه‌سازی تکنیک‌های موجود در حوزه‌های مرتبط به سرعت در حال پیشرفت هستند. سیستم بینایی کامپیووتر ممکن است برای بررسی تصاویر در ابتدا به الگوریتم‌های پردازش تصویر نیاز داشته باشد و اعمالی را برای بهبود عملکرد خود بر روی تصاویر انجام دهد. بینایی کامپیووتر دارای کاربردهای متنوعی است و یک فناوری در حال توسعه محسوب می‌شود. با توجه به پایین بودن میزان خطای سرعت بالا، هزینه‌های نگهداری پایین و عدم نیاز به حضور مداوم نیروی انسانی، تعداد زیادی از صاحبان صنایع و کارخانه‌ها به سرعت به سمت بینایی کامپیووتر روی آورده‌اند. امروزه در کارخانه‌های پیشرفته از سیستم بینایی کامپیووتر به منظور مدیریت سیستم‌های ایمنی موجود در محیط‌های صنعتی، کنترل و بررسی مواد اولیه، میزان کیفیت محصول و ارتقا آن و کنترل موجودی انبار استفاده می‌شود.

از کاربردهای امنیتی بینایی کامپیووتر در زندگی روزمره سیستم تشخیص هویت بر اساس مردمک چشم انسان است. دانشمندان به این نتیجه رسیده‌اند که هر فردی دارای الگوهای منحصر به فردی در مردمک چشم خود است که او را از افراد دیگر متمایز می‌کند و همین امر در شناسایی افراد حائز اهمیت است.

از دیگر مسائل شمارش تعداد اجسام که نقش بینایی کامپیووتر در حل آن حائز اهمیت است، می‌توان به شمارش خودروهای داخل پارکینگ که شکل 2 نمایی از آن را نشان می‌دهد، اشاره کرد. سیستم‌های مجهرز به بینایی کامپیووتر در پارکینگ‌های هوشمند این امکان را فراهم می‌آورند که ظرفیت موجود را کنترل کنند و به طور دقیق خودرو ورودی جدید را به مکان پارک هدایت کند یا

از ورود خودرو به پارکینگ به علت نبود جای پارک خودداری کند. از طرفی این تکنولوژی با دارا بودن قابلیت تشخیص پلاک‌های وسایل نقلیه و ثبت تخلفات رانندگی به راهنمایی رانندگی در کنترل و مدیریت ترافیک کمک شایانی می‌کند.

همراه با تغییر فرهنگی مستمر در ارتباطات، جامعه نیز روش‌های ارتباطی خود را از حالت‌های کلامی و مبتنی بر متن به حالت‌های متن با تصویر گسترش داده است. افراد بسیاری از فعالیت‌های خود مثل رزرو هتل، سفارش غذا و خرید کالا را به صورت غیرحضوری و از طریق اینترنت انجام می‌دهند و به تصاویر ارائه شده توجه می‌کنند و سپس توضیحات و نظرات را بررسی می‌کنند. انتخاب تصویر باکیفیت از محصول حائز اهمیت است و در موقیت خدمات و جذب مشتری مؤثر است. شرکت‌ها و افرادی که دارای روابط تجاری هستند از این استفاده همگانی مردم بهره برده و از آن برای تقویت روابط خود و کسب و کارشان استفاده می‌کنند.



شکل ۲: نمایی از شناسایی خودروها در پارکینگ [۷۴]

### شبکه عصبی

شبکه عصبی<sup>۱</sup> یکی از اصلی‌ترین و مهم‌ترین موضوع در مبحث یادگیری ماشین و هوش مصنوعی است. شبکه‌های عصبی قابلیت این را دارند که بدون دریافت دستورالعمل از جانب خبره، بر اساس مسئله پیش‌رو، مدلی بر روی دادگان آموزشی تولید کرده و عملکرد خود را بهبود بخشنند.

شبکه عصبی مانند کودکان با مثال یاد می‌گیرد و برای انجام وظیفه‌های مشخص مانند

<sup>۱</sup>Neural Network (NN)

شناسایی الگوها و دسته‌بندی اطلاعات، در طول یک پروسه یادگیری تنظیم می‌شود. در سیستم‌های زیستی، یادگیری با تنظیماتی در اتصالات سیناپسی که بین اعصاب قرار دارد همراه است و این اساس پایه‌گذاری شبکه‌های عصبی است.

یک شبکه عصبی روشی برای تحلیل اطلاعات است که از سیستم عصبی انسان الهام گرفته و سعی دارد از روشی که نورون‌های مغز انسان پردازش و ورودی‌های حواس انسان را درک می‌کند، تقليید کند و به پردازش اطلاعات و دسته‌بندی اطلاعات بپردازد. این سیستم پردازشی از تعداد زیادی عناصر به هم پیوسته به نام نورون‌ها<sup>۱</sup> تشکیل شده که برای حل یک مسئله به‌طور هماهنگ با یکدیگر عمل می‌کنند.

### یادگیری عمیق

یادگیری عمیق یک تکنیک یادگیری ماشین است. فرایند آموزش در روش یادگیری عمیق به این صورت است که ویژگی‌های مربوطه به‌طور خودکار از تصاویر استخراج می‌شوند. طی سال‌های اخیر، یادگیری عمیق مبتنی بر شبکه عصبی در حوزه بینایی کامپیوتر به‌طور وسیع مورد مطالعه قرار گرفته است. در روش‌های یادگیری عمیق از معماری شبکه عصبی عمیق استفاده می‌شود. اصطلاح عمیق به تعداد لایه‌های پنهان در شبکه عصبی اشاره دارد. شبکه‌های عصبی سنتی فقط شامل ۳-۲ لایه مخفی هستند، در حالی که شبکه‌های عمیق می‌توانند تا ۱۵۰ لایه داشته باشند. در لایه‌های اولیه شناسایی لبه‌ها، در لایه‌های میانی شناسایی اشکال و در لایه‌های نهایی تشخیص ویژگی‌های خاص مدنظر صورت می‌گیرد.

یادگیری عمیق، یادگیری انتهایی به انتها را انجام می‌دهد، به این معنی که اگر وظیفه‌ی یک شبکه طبقه‌بندی باشد، آن شبکه برای هر داده‌های خام ورودی طبقه متناظر را پیش‌بینی کرده و خروجی می‌دهد. انجام این کار توسط شبکه مبتنی بر یادگیری عمیق به‌طور خودکار خواهد بود. هدف از روش یادگیری عمیق، یادگیری سلسله مراتبی ویژگی‌ها با استفاده از استخراج ویژگی‌های سطوح بالاتر است که از ترکیب ویژگی‌های سطح پایین‌تر تشکیل شده است.

<sup>1</sup>Neurons

تفاوت اصلی یادگیری ماشین و یادگیری عمیق در استخراج ویژگی است. ویژگی‌های شاخص تصویر در یادگیری ماشین، به صورت دستی محاسبه می‌شود. استخراج ویژگی در یادگیری عمیق به این صورت است که در مدت زمان کافی تعداد دادگان عظیمی با برچسب در اختیار الگوریتم یادگیری قرار می‌گیرد تا الگوها و ویژگی‌های شاخص و معین تصویر را یاد بگیرد و بازنمایی مطلوبی از آن‌ها ارائه دهد. در یادگیری عمیق رایانه یاد می‌گیرد که کارهای طبقه‌بندی را مستقیم از طریق تصاویر، متن یا صدا انجام دهد. به عنوان مثال، توسعه خودرو خودران به میلیون‌ها تصویر و هزاران ساعت فیلم نیاز دارد تا به کمک یادگیری عمیق بتواند علامت توقف را تشخیص دهد یا عابر پیاده را از چراغ راهنمایی تمايز دهد.

یادگیری عمیق یک الگوریتم خودسازگار است به این معنی که مدل‌های آموزش دیده به روش یادگیری عمیق الگوهای بهتری را با تجربه داده‌های جدید کشف می‌کنند و به بهبود خود ادامه می‌دهند. مدل آموزش دیده شده توسط یادگیری عمیق در حد خبره‌های انسانی و یا حتی بهتر از آن عمل می‌کند.

هدف از آموزش شبکه‌های عمیق یافتن پارامترهای بهینه با استفاده از داده‌های آموزشی است. شبکه‌های عمیق دارای تعداد زیادی پارامتر ناشناخته هستند. برای یافتن دقیق همه پارامترهای ناشناخته، به داده‌های زیادی نیاز است. اگر داده‌های کمی موجود باشد، برای اکثر پارامترها فقط مقادیر تقریبی به دست می‌آید و مدل به شکل قابل تعمیمی یاد نمی‌گیرد یعنی برای داده جدید تشخیص اشتباه می‌دهد که مطلوب نیست. در واقع برای شبکه‌های عمیق باید به این نکته مهم که هر چه داده‌های بیشتری برای آموزش در دسترس باشد، یادگیری مدل بهتر و مؤثرتر خواهد بود؛ اما فراهم آوردن چنین مجموعه عظیمی از داده‌های برچسب‌دار برای آموزش شبکه، دشوار است. از طرف دیگر برای پردازش حجم عظیمی از دادگان و محاسبه مقادیر پارامترها به پردازنده‌های گرافیکی دارای معماری موازی با قدرت محاسبات و پردازش بالا نیاز است که لازمه صرف هزینه زیادی است. به عنوان راه حل برای موانع مطرح شده روش یادگیری انتقالی مطرح گردید. در ادامه این روش بیان می‌شود.

## یادگیری انتقالی

روش یادگیری انتقالی از دانش مدلی که از قبل برای وظیفه‌ای دیگر و توسط حجم زیادی از داده‌ها مانند ImageNet که شامل بیش از ۱۴ میلیون تصویر متعلق به بیش از ۲۰۰۰ کلاس است طی صدها ساعت بر بستر پردازنده‌های گرافیکی قدرتمند آموزش دیده برای حل وظیفه‌ای متفاوت استفاده می‌کند. در این روش به جای این که از ابتدا مدلی آموزش ببیند، از این مدل‌ها به عنوان نقطه شروع فرآیند آموزش استفاده می‌شود. به طوری که ویژگی‌هایی که در حین آموزش روی کار قبلی آموخته شده است، برای کار جدید و متفاوت از وظیفه پیشین به کار می‌گیرد. به بیان دیگر مقادیری که برای پارامترهای مدل در فرآیند آموزش تعیین شده است در کار جدید استفاده می‌شود.

ایده‌ی 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 شود؛ به این معنی که یادگیری شبکه از صفر شروع نمی‌شود بلکه با الگوهای یاد گرفته شده که در حین حل کردن وظیفه‌ای دیگر به دست آمده است، شروع می‌شود. این روش، این مزیت را دارد که به داده‌های کمتری نیاز است. بنابراین زمان محاسبه کاهش می‌یابد. گروه‌های تحقیقاتی، مدل‌هایی را که برای مسابقه‌هایی مانند ILSVRC آموزش داده‌اند به صورت منبع باز به اشتراک می‌گذارند که می‌توان آن‌ها را برای مسائل مختلف انتخاب کرد.

مطابق شکل ۳ مراحل انجام یادگیری انتقالی را می‌توان به صورت زیر بیان کرد.

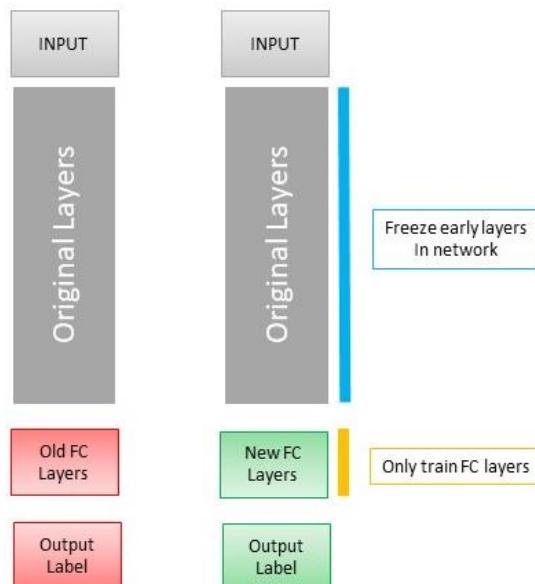
(۱) ابتدا مجموعه جدیدی از لایه‌های کاملاً متصل با مقداردهی اولیه تصادفی جایگزین مجموعه‌ای از لایه‌های کاملاً متصل از یک شبکه عصبی کانولوشنال از پیش آموزش دیده می‌شود.

(۲) تمام لایه‌ها به جز سر شبکه منجمد<sup>۱</sup> می‌شوند، به این معنی که پارامترهای آن لایه‌ها به روز نمی‌شوند و عبور به عقب در مرحله پس انتشار به آن‌ها نمی‌رسد.

<sup>۱</sup>ImageNet Large Scale Visual Recognition Challenge

<sup>۲</sup>Freeze

سپس شبکه با استفاده از نرخ یادگیری بسیار کوچک آموزش می‌بیند تا مجموعه جدید لایه‌های کاملاً متصل بتوانند الگوهایی را از لایه‌های کانولوشنال قبل‌آموخته شده در شبکه یاد بگیرند.

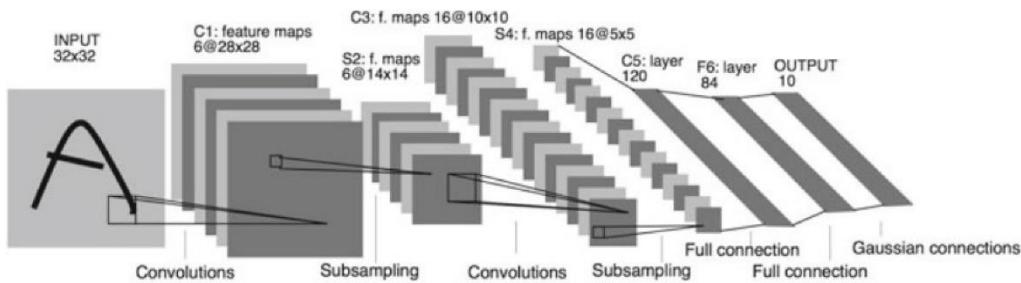


شکل ۳:: مراحل یادگیری انتقالی

در ادامه معماری‌های مهم CNN که در گذر زمان ارائه شده‌اند معرفی و تاریخچه مختصری از آن بیان می‌شود.

### معماری LeNet

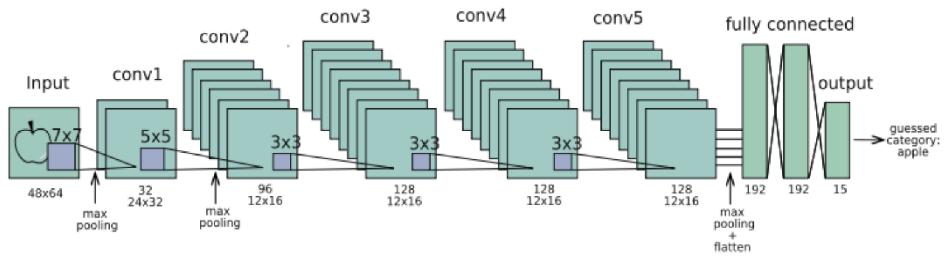
اولین معماری CNN بود که در سال ۱۹۹۸ معرفی شد [۷۵]. ساختار از سه لایه کانولوشن و دو لایه کاملاً متصل تشکیل شده است. معماری LeNet در شکل ۴ نشان داده شده است. شبکه شامل پنج لایه با پارامترهای قابل یادگیری است و ترکیب و میانگین ادغام و سه مجموعه از لایه‌های کانولوشن را ترکیب می‌کند. دو لایه کاملاً متصل پس از فرآیند کانولوشن و ادغام وجود دارد. در پایان، یک طبقه‌بندی‌کننده softmax تصاویر را طبقه‌بندی می‌کند.



شکل ۴: معماری معماري شبکه LeNet [۷۵]

### معماری AlexNet

در سال ۲۰۱۲ در خلال چالش ImageNet پیشنهاد شد و برنده مسابقه شد [۷۶]. معماری پیشنهادی با استفاده از لایه‌های کانولوشن، حداکثر لایه‌های ادغام، افزایش داده، حذف، فعال‌سازی ReLU و SGD، خط را از ۲۶ درصد به  $\frac{1}{3}$  درصد کاهش داد. AlexNet با ۶۰ میلیون پارامتر دارای هشت لایه، پنج لایه کانولوشن و سه لایه کاملاً متصل است. شکل ۵ معماری شبکه کانولوشنال AlexNet را نشان می‌دهد که توسط پتینو و همکاران پیشنهاد شده است. شبکه AlexNet به منظور طبقه‌بندی میوه‌های استوایی با ۲۶۳۳ تصویر متنوع از میوه‌ها به ۱۵ دسته استفاده گردید [۷۷]. نویسنده‌گان [۷۸] از AlexNet برای آموزش مجموعه داده‌های مختلف متشکل از تصاویر سبزیجات استفاده کردند. بر اساس آزمایش، میزان دقیقت در مقایسه با روش SVM با ۸۰/۵ درصد، به ۹۲/۱ درصد رسید.



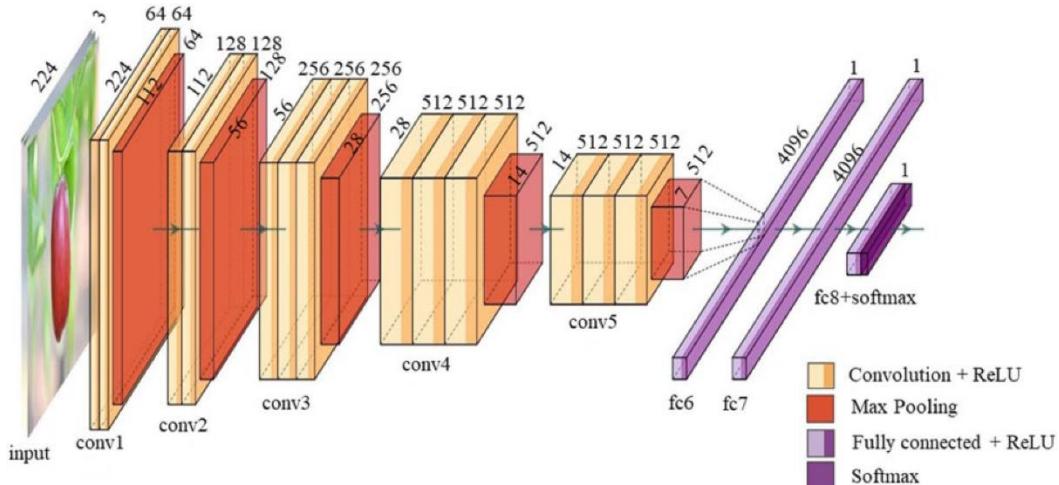
[۷۷] معماری AlexNet

## معماری VGGNet

شبکه VGG جزء اولین شبکه‌های عصبی کانولوشنال است که برای طبقه‌بندی بیش از ۱۴ میلیون تصویر ImageNet که شامل ۱۰۰۰ کلاس است طراحی شد و توانست دقیق‌تر باشد در مسابقه محبوب‌ترین طبقه‌بندی کننده تصویر در نظر گرفته می‌شد، با تغییر اندازه فیلتر هسته پیشنهاد شد [۴۰]. هدف از طراحی VGG بهبود زمان آموزش و کاهش تعداد پارامترها بود.

بعد از هر یک از لایه‌های کاملاً متصل در معماری شبکه VGG یک تابع فعال‌ساز ReLU اعمال می‌شود. استفاده از توابع فعال‌سازی دیگری نظیر  $\tanh$  نیز در شبکه‌های عصبی عمیق کانولوشنال متداول است؛ اما تابع فعال‌سازی  $\tanh$ ، سرعت همگرایی کمتری نسبت به ReLU دارد و از طرفی در شبکه‌های عصبی عمیق‌تر، با احتمال بیشتری دچار بیش برآش خواهد شد. به دلیل سادگی و تعداد لایه‌های کم این شبکه هنوز هم یکی از پرکاربردترین معماری‌های شناسایی تصویر محسوب می‌شود. شبکه VGG انواع زیادی دارد که در ادامه تنها به معماری ۱۶-VGG اشاره می‌شود. معماری ۱۶-VGG از پنج لایه کانولوشن و سه لایه کاملاً متصل حاوی ۱۳۸ میلیون پارامتر تشکیل شده است [۷۹]. دو لایه اول کاملاً متصل شامل ۴۰۹۶ گره است و لایه خروجی VGGNet با توجه به این که برای طبقه‌بندی تصاویر ImageNet به ۱۰۰۰ کلاس طراحی شده است، شامل ۱۰۰۰ نورون است. شکل ۶ معماری VGG16 که توسط منگ و همکاران برای طبقه‌بندی عناب پیشنهاد شده است را نشان می‌دهد [۸۰]. VGG16 برخلاف AlexNet، دارای یک شبکه عمیق‌تر و ساختار یکنواخت است که از ۱۶ لایه قابل آموزش شامل ۱۳ لایه کانولوشن و سه لایه کاملاً متصل تشکیل شده است. این شبکه یک تصویر رنگی به اندازه ۲۲۴×۲۲۴ پیکسل را دریافت می‌کند.

سپس تصویر ورودی از لایه‌های کانولوشن عبور داده می‌شود تا اطلاعات برجسته از آن استخراج شود. لایه‌های کانولوشن در این شبکه از فیلترهایی با ابعاد  $3 \times 3$  که کوچک‌ترین اندازه‌ی ممکن برای فیلتر محاسبه می‌شود، استفاده از فیلترهایی با ابعاد کوچک در ساختار شبکه پارامترها را کاهش می‌دهد که در نتیجه سرعت یادگیری شبکه افزایش می‌یابد.



شکل ۶: معماری شبکه VGG-16 برای طبقه‌بندی عناب [۸۰]

## معماری ResNet

معماری ResNet توسط شرکت مایکروسافت معرفی شد [۵۲]. نام این معماری برگرفته از Network به معنای شبکه باقیمانده است. شبکه‌های یادگیری عمیقی مانند AlexNet و VGGNet، اغلب لایه‌های کانولوشنال و سپس لایه‌های کاملاً متصل بدون هیچ‌گونه اتصال میان برای طبقه‌بندی دارند. به همین دلیل آن‌ها را شبکه‌های ساده<sup>۱</sup> می‌نامند. وقتی شبکه‌ی ساده عمیق‌تر می‌شوند، مشکل محوشدگی گرادیان<sup>۲</sup> یا انفجار گرادیان<sup>۳</sup> رخ می‌دهد و منجر به رفتار ناپایدار شبکه می‌شود. به همین دلیل تنها با اضافه کردن چند لایه به شبکه و عمیق‌تر کردن آن، عملکرد شبکه را نمی‌توان بهبود داد.

برخی توابع فعال‌ساز، مانند تابع سیگموید<sup>۴</sup>، مقادیر ورودی با مقیاس بزرگ را در یک بازه‌ی

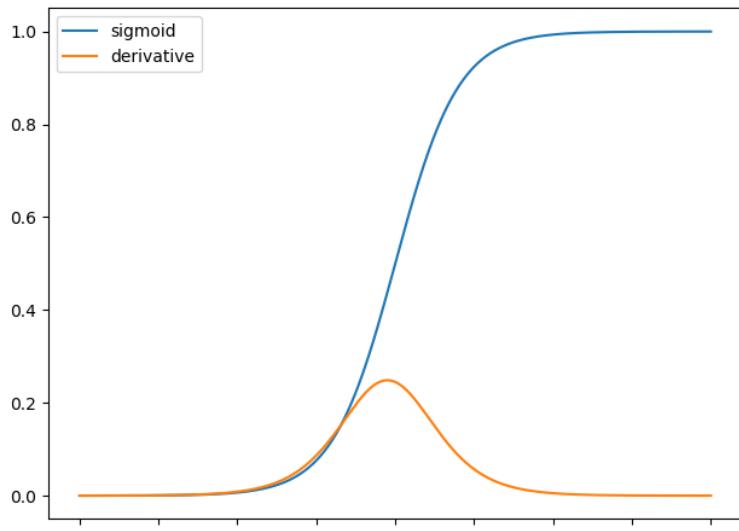
<sup>۱</sup>Plain Networks

<sup>۲</sup>Vanishing Gradient

<sup>۳</sup>Exploding Gradient

<sup>۴</sup>Sigmoid

کوچک میان صفر و ۱ قرار می‌دهند؛ بنابراین زمانی که یک تغییر بسیار بزرگ در مقدار ورودی تابع اتفاق می‌افتد، خروجی تابع تنها مقدار کمی تغییر می‌کند؛ این یعنی مقدار مشتق آن خیلی کوچک می‌شود. برای مثال، در شکل ۷ تابع Sigmoid را به همراه مشتق آن نشان می‌دهد. زمانی که مقدار ورودی تابع بزرگ یا کوچک می‌شود، مقدار مشتق آن به صفر نزدیک می‌شود.

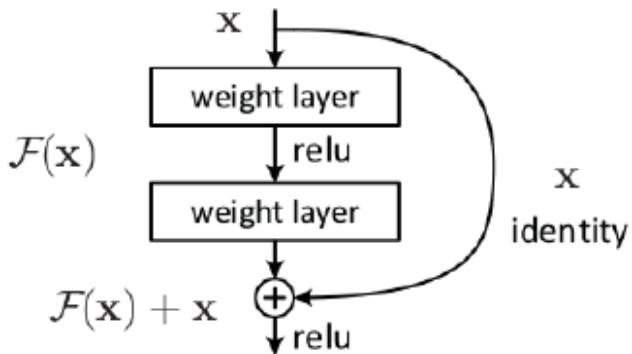


شکل ۷: نمایی از تابع sigmoid و مشتق آن

برای رفع مشکل محوشدگی گرادیان شبکه‌ی ResNet معرفی شد. در این شبکه، بین لایه‌ها ارتباطاتی در کنار ساختار کانولوشنال پیاده‌سازی شده است. به این نحو که ورودی‌های لایه قبلی بدون واسطه در اختیار لایه بعدی قرار می‌گیرد تا بتوان شبکه را عمیق‌تر کرد و روند آموزش آن را سرعت بخشد. به این ارتباطات skip connection و به ساختار حاصل از آن ماضول باقیمانده<sup>۱</sup> گفته می‌شود. شکل ۸ نمایی از ماضول باقیمانده را نشان می‌دهد. عملکرد آن به این نحو است که می‌تواند یک مجموعه عملیات را بر روی داده ورودی انجام دهد یا کلاً نادیده بگیرد. در واقع اطلاعات از بین رفته احتمالی در مرحله کانولوشن را به لایه بعدی انتقال می‌دهد.

---

<sup>۱</sup>Residual Network



شکل ۸: نمایی از مازول باقیمانده [۵۲] ساختار اصلی معماری ResNet از چندین مازول باقیمانده کنار هم تشکیل شده است. ResNet از معماری VGGNet الهام گرفته شده با این تفاوت که به شبکه VGG اتصالات میان بر اضافه شده است.

این شبکه در نسخه های مختلفی از جمله ResNet152, ResNet50, ResNet34 ارائه شده است. ResNet152 ۱۵۲ لایه دارد و ۸ برابر عمیق تر از VGG16 است هر قدر شبکه عمیق تر می شود، پیچیدگی زمانی نیز افزایش می یابد. برای رفع این مشکل، به اول و آخر هر لایه کانولوشن یک لایه کانولوشن  $1 \times 1$  اضافه شد. تکنیک کانولوشن  $1 \times 1$  در شبکه GoogleNet استفاده شده است و نشان می دهد کانولوشن های  $1 \times 1$  می توانند تعداد پارامتر های شبکه را کاهش دهند و کارایی آن را کاهش ندهند. این نوع ساختار هنگام افزایش لایه ها مشکلی در کاهش عملکرد نخواهد داشت. استفاده از تکنیک کانولوشن  $1 \times 1$  در معماری ResNet152 باعث شده است که به مرتب پیچیدگی کمتری نسبت به VGG16 داشته باشد.

## **Abstract**

In recent years, multimedia content, especially digital images, has played an important role in human daily life due to its visual nature and impact. The most effective feature in the impact of images is their quality. Given the importance of image quality in applied fields, its evaluation has been considered and requires the need to evaluate image quality automatically. Evaluating image quality without knowing the type of damage applied to the image and without access to the reference image is one of the challenges in the field of digital image processing. In this context, comparing a distorted image with a reference image is not possible. Given the development and success of the deep learning field in providing powerful methods for solving problems, this research used a deep learning approach and convolutional neural network reconfiguration to evaluate image quality automatically. The proposed model is based on the Fastai architecture and using transfer learning. This research presents a model for the image quality assessment problem, which is a regression problem, by reconfiguring the ResNet18 convolutional neural network that has been previously trained on the ImageNet image for the classification problem. This model is trained on the TID2013 image dataset consisting of 3000 images. The performance of the proposed model was compared with other previous models presented in this problem. The performance of the model based on the assessed image quality with the root mean square error criterion is 0.56, which is better than other previous models and is more

reliable for assessing the quality of an image. The results of this research indicate that the proposed model can be used to more reliably assess the quality of an image automatically in its application areas such as image distortion detection, image enhancement, and removal of destructive factors.

**Keywords:** Image quality assessment, deep learning, TID2013 database.



Department of Computer Engineering

**A Thesis Submitted in Partial Fulfillment of the Requirements for the  
Degree of Master of Science in Computer Engineering-Artificial  
Intelligence and Robotics**

**Automatic Document Image Quality Assessment  
using Deep Neural Network**

**By:**

Reyhaneh Ghafouripour

Evaluated and Approved by Thesis Committee, in Winter 2025

1. Dr. Ali Mohammad Latif, Assoc. Professor (Supervisor)
2. Dr. Mahdi Yazdian Dehkordi, Assist. Professor (Advisor)
3. Full name, Assoc. Professor (Examiner)
4. Full name, Assoc. Professor (Examiner)
5. Full name, Assoc. Professor (Examiner)