

رنگ آمیزی خود کار تصاویر خاکستری به کمک شبکه های زایای رقابتی

محمد مهدی حاجی اسمعیلی و غلامعلی منتظر*

گروه مهندسی فناوری اطلاعات، دانشگاه تربیت مدرس، تهران، ایران

چکیده

رنگ آمیزی تصاویر خاکستری یکی از مسائل مهم در حوزه بازیابی اطلاعات تصویری محسوب می شود. البته میزان موفقیت روش های خود کار در این حوزه در برابر عملکرد گرافیک ها و ویراستاران تصویر ناچیز بوده است. از بین رفتن اطلاعات رنگ ها و دست یابی به قسمت محدودی از اطلاعات اولیه تصاویر، این موضوع را به چالشی منحصر به فرد تبدیل می کند؛ چون هدف اصلی در رنگ آمیزی تصاویر خاکستری، پیدا کردن رنگ اصلی و واقعی تصویر نیست؛ بلکه تلاش بر این است تا نوعی رنگ آمیزی که از نظر انسان ها «واقعی» به نظر می رسد، برای تصویر ارائه شود. در این مقاله سامانه ای برای رنگ آمیزی تصاویر خاکستری اشخاص و انسان ها ارائه شده است. در این سامانه که نیاز به پیش پردازش تصاویر یا مؤلفه های پردازشی مستقل ندارد، از یک شبکه عصبی عمیق برای رنگ آمیزی تصاویر افراد و از شبکه ای دیگر برای بررسی باز خورد عملکرد رنگ آمیزی استفاده می شود. نتایج حاصل از پژوهش، عملکرد کیفی بهتری را در رنگ آمیزی تصاویر افراد در مقایسه با روش های مرسوم در این عرصه نشان می دهد که در مدت زمان آموزش، به مراتب کمتری به دست آمده است. کاربرد این سامانه نه تنها در رنگ آمیزی تصاویر افراد، بلکه در رنگ آمیزی پرتوهای غیر انسانی و آثار هنری نیز کاربرد دارد و به کمک آن می توان رنگ آمیزی مناسبی برای اثرهایی که ممکن است، ساعت ها زمان ببرند، در کسری از ثانیه انجام داد.

واژگان کلیدی: رنگ آمیزی، تصاویر خاکستری، شبکه های عصبی عمیق، شبکه های زایای رقابتی، بازیابی اطلاعات تصویری

Automatic Colorization of Grayscale Images Using Generative Adversarial Networks

Mohammad Mahdi Haji-Esmaili & Gholam Ali Montazer*

Information Technology Engineering, Tarbiat Modares University, Tehran, Iran

Abstract

Automatic colorization of gray scale images poses a unique challenge in Information Retrieval. The goal of this field is to colorize images which have lost some color channels (such as the RGB channels or the AB channels in the LAB color space) while only having the brightness channel available, which is usually the case in a vast array of old photos and portraits. Having the ability to colorize such images would give us a multitude of possibilities ranging from colorizing old and historic images to providing alternate colorizations for real images or artistic creations. Be that as it may, the progress in this field is trivial compared to what the professionals are able to do using special-purpose applications such as Photoshop or GIMP. On the other hand, losing the information stored in color channels and having only access to the primary brightness channel, makes this problem a unique challenge, since the main aim of automatic colorization is not to find the image's "real" color but to colorize it in such a way that makes it "seem real" as the original color information is lost forever and the only way to colorize it, is to provide a somewhat "proper" estimation. In this research we propose a model to automatically colorize gray human portraits. We start by reviewing the methods used for the task of image colorization and provide an explanation as to why most of them collapse to a situation known as "Averaging". To counteract this effect, we design our end-to-end model with two

* Corresponding author

* نویسنده عهده دار مکاتبات

separate deep neural networks forming a Generative Adversarial Network (GAN), one to colorize the images and the other to evaluate the colorization of the first network and guide it towards the proper distribution. The results show improvements over other proposed methods in this field especially in the case of colorizing human portraits along faster train times. This method not only works on real human portraits but also on non-human and artistic portraits that can be leveraged to colorize hand-drawn images some of which may take minutes up to hours by hand.

Keywords: colorization, gray scale images, deep neural networks, generative adversarial networks, image information retrieval

نمایش داده شده است:



(شکل-1): تفاوت شدت رنگ خاکستری و پیوستگی رنگ‌ها برای

انتخاب رنگ «قابل قبول»

(Figure-1): The difference between grayscale intensity and color diversity for choosing the "proper color"

در تصاویر شکل (1)، تفاوت شدت روشنایی موجب به وجود آمدن جایگشت‌های مختلفی برای رنگ‌های قابل قبول این پیراهن‌ها می‌شود؛ به طوری که با از دست رفتن اطلاعات اصلی رنگ درون تصاویر، به طور تقریبی هر ترکیبی از رنگ‌های متفاوت را می‌توان برای این لباس‌ها در نظر گرفت تا جایی که حتی در نقاطی که پیوستگی رنگی یکسان است، می‌توان از لکه‌های رنگی متفاوت استفاده کرد. این مشکل که ارتباط نزدیکی با «سرریز شدن رنگ¹» دارد خود را در مواردی همچون پوست‌هایی که رنگ لباس به درون آن‌ها نفوذ کرده تا لباس‌هایی که رنگ یک‌دستی ندارند، نشان می‌دهد. یکی دیگر از چالش‌های مهم در این حوزه، «رنگ‌آمیزی معنایی²» اشیای موجود در تصویر است که بحث اصلی آن رنگ‌آمیزی بر اساس مفاهیم معنایی است که به طور تقریبی همیشه برای اشیاء و رنگ آن‌ها صادق است. به عنوان مثال کسی که لباس آستین کوتاه پوشیده است، نباید دستانی با رنگ‌های متفاوت داشته باشد، یا دو عدد توپ بازی همسان که در یک اتاق بر

۱- مقدمه

در حال حاضر با حجم عظیمی از تصاویر قدیمی و سیاه و سفید که سال‌ها پیش گرفته شده‌اند، روبه‌رو هستیم و راه کار مستقیمی برای رنگ‌آمیزی آنها وجود ندارد. رنگ‌آمیزی دستی نیازمند دخالت مستقیم کاربر و فرایندی زمان‌بر است که در مواردی همچون بررسی تصاویر همسان (برای انتخاب رنگ)، قطعه‌بندی تصویر (برای اعمال رنگ) و مقایسه چشمی با تصاویر همسان (برای بررسی تناسب رنگ) نمود پیدا می‌کند [1]. رنگ‌آمیزی دستی با استفاده از ابزارهایی همچون Adobe Photoshop یا Arkvis Coloriage انجام می‌شود که نیاز به تخصص در کاربرد ابزارهای یادشده دارد. هم‌اکنون گروه‌های مباحثه و شبکه‌های اجتماعی متعددی شکل گرفته‌اند که از کاربران برای رنگ‌آمیزی تصاویر به صورت دستی (که اغلب از آلبوم‌های خانوادگی هستند) کمک می‌گیرند که در بین آن‌ها می‌توان به گروه‌هایی همچون r/ColorizationRequests و یا r/Colorization اشاره کرد [2].

مسئله اصلی در این پژوهش، رنگ‌آمیزی قابل قبول برای تصاویر خاکستری از پرتوهای انسانی در ابعادی مشخص است. «قابل قبول» بودن رنگ‌آمیزی به عناصر مختلفی بستگی دارد، به عنوان مثال در حوزه رنگ‌آمیزی پوست صورت، بازه‌ای از رنگ‌های محدود که اغلب در طیفی میان سفید، قهوه‌ای و سیاه هستند «قابل قبول» در نظر گرفته می‌شوند در حالی که رنگ‌آمیزی پوست انسان‌ها با رنگ‌هایی همچون سبز، آبی یا فسفری به هیچ وجه قابل قبول محسوب نمی‌شود [3]. از طرفی شدت روشنایی و پیوستگی توزیع آن در سطوحی همچون لباس یا پوست، عنصر دیگری از عناصر تأثیرگذار در مقبولیت رنگ‌آمیزی محسوب می‌شود؛ در لباس‌هایی که شدت روشنایی پیوسته و مشخصی دارند، نمی‌توان انتظار داشت که قسمتی از لباس به رنگ آبی و قطعات و تکه‌هایی از آن قرمز باشد و حتی اگر سایه‌ای بر روی لباس افتاده و میزان روشنایی آن را در قطعاتی دچار تغییر کند، باز هم انتظار می‌رود که رنگ اصلی لباس در این قطعات فقط با شدتی متفاوت حضور داشته باشد. در شکل (1) نمونه‌ای از چالش «مقبولیت»

¹ Color Bleeding

² Semantic Colorization

میانگینی از تصاویر پوست اغلب منجر به رنگ سوبیایی می شود که در ستون سمت راست شکل (۲) دیده می شود. این موضوع در مورد رنگ قرمز لب های افراد نیز صادق است که مدل به ناچار با میانگین گیری از آن باز هم به رنگ سوبیا برای لب ها می رسد. این چالش نه تنها در رنگ آمیزی صورت ها و افراد وجود دارد، بلکه در رنگ آمیزی دنیای اطراف نیز خود را نشان می دهد. در شکل (۳) نمونه هایی از رخداد میانگین گیری در رنگ آمیزی تصاویری با دامنه های متفاوت نمایش داده شده است.



(شکل-۲): تأثیر تابع هدف MSE بر رنگ آمیزی صورت
(Figure-2): The effect of MSE on colorizing portraits [5]



(شکل-۳): میانگین گیری در رنگ آمیزی تصاویر
(Figure-3): Averaging problem in image colorization [6]

در تصویر (الف) قسمت اعظمی از تصویر یا رنگ آمیزی نشده و یا دارای رنگ سوبیا است و از طرفی رنگ آمیزی ماشین به ناچار تبدیل به رنگ قرمز کدر شده است. در تصویر (ج) سامانه به نتیجه رسیده که برای کاهش خطای کل تصویر، با میانگین گیری به خطای کمتری می رسد تا اینکه بخواهد اشیای درون تصویر را به صورتی جداگانه رنگ آمیزی کند که به احتمال با خطای بیشتری همراه خواهد بود.

با مقایسه هیستوگرام تصاویر رنگ آمیزی شده، می توان به تأثیر میانگین گیری بر روی فضای رنگ تصاویر پی برد. در

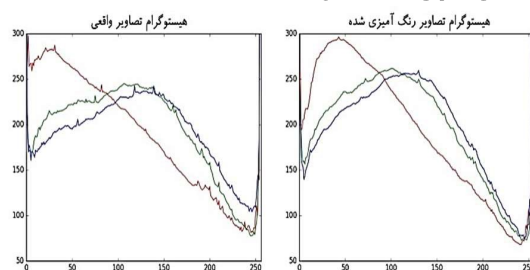
روی زمین افتاده اند، نباید ساختار رنگی متفاوتی داشته باشند؛ یعنی برای مثال یکی به صورت رنگین کمانی رنگ شود و دیگری با همان ساختار به صورت قهوه ای یکدست.

عموم روش هایی که برای حل این مسأله استفاده شده اند، دارای مشکلی اساسی هستند که خود را در رنگ آمیزی تصاویر با رنگ های مُرده و اغلب به رنگ سوبیا نشان می دهد (رنگ سوبیا، رنگی قهوه ای با درجه کمی از قرمزی است که به دلیل فعل و انفعالات شیمیایی، بخش اعظمی از تصاویر شکار شده با دوربین های قدیمی در طول زمان به این رنگ در می آیند) [4]. رنگ آمیزی، مسأله ای از نوع رگرسیون است و اغلب روش های به کار گرفته شده در آن از تابع هدفی همچون میانگین مربع خطا (MSE) استفاده می کنند. با توجه به اینکه هدف نهایی در این مسأله، رنگ آمیزی تصویر خاکستری، به نحوی است که بیشترین شباهت را با تصویر رنگی واقعی داشته باشد، به کارگیری چنین تابع خطایی مناسب به نظر می رسد. البته این انتخاب، عملکرد بهینه سازی را به صورت مستقیم تحت تأثیر قرار می دهد و سامانه را به سمتی سوق می دهد تا با انتخاب رنگ هایی مناسب برای کل تصویر، میزان خطا را کاهش دهد [4]. این جهت گیری، سامانه را وادار به انتخاب رنگ های میانه ای برای بسیاری از اشیای درون تصویر می کند. به عنوان نمونه ای مناسب از این عملکرد می توان به رنگ آمیزی آسمان، دریا یا چمن اشاره کرد؛ زیرا این موارد در بیش تر حالات دارای رنگ های یکسانی هستند، به همین دلیل سامانه با انتخاب رنگی در حوزه های آبی و سبز، آن ها را رنگ آمیزی خواهد کرد. با وجود عملکرد مثبت سامانه در رنگ آمیزی مفاهیمی با رنگ ثابت، این موضوع تأثیری منفی در رنگ آمیزی اشیایی که دارای تنوع رنگ زیاد هستند، دارد. به عنوان مثال، سامانه با دیدن یک ماشین در تصویر، نمی داند که از چه رنگی برای آن استفاده کند؛ چون در تصاویری متعدد ماشین هایی با رنگ قرمز، نارنجی، سرمه ای، مشکی، نقره ای و سفید دیده است. در این موارد، تابع خطایی همچون MSE، مدل را مجاب به انتخاب رنگی می کند که میانگینی از رنگ هایی است که برای ماشین دیده است. این مشکل که با نام «میانگین گیری» نیز شناخته می شود، موجب ایجاد تصاویری با رنگ های مُرده و اغلب سوبیا می شود که ناشی از میانگین گیری بیش تر رنگ هایی است که مدل از اشیای دیده است. در شکل (۲) تعدادی از این رنگ آمیزی ها نمایش داده شده اند.

¹ Sepia

² Mean Squared Error

شکل (۴) نمایی از هیستوگرام بیش از سه هزار تصویر نمایش داده شده است. نرم شدن هیستوگرام‌ها و کاهش واریانس در کانال‌های رنگ مختلف، نشان از بی‌تفاوتی مدل به رنگ آمیزی متنوع اشیای درون تصویر و میانگین‌گیری آن از رنگ‌ها برای کاهش میزان خطای کل است.



(شکل-۴): هیستوگرام رنگی تصاویر افراد و صورت
(Figure-4): Color image histograms of faces and bodies

با توجه به نکات بالا، این مقاله شامل بخش‌های زیر است: در بخش دو نگاهی مختصر به ساختار و عملکرد شبکه‌های زیای رقابتی خواهیم داشت و سپس در بخش سه به نتایج پژوهش‌های در زمینه رنگ‌آمیزی و محدودیت‌های هر روش خواهیم پرداخت. در بخش چهار به بررسی روش پیشنهادی می‌پردازیم و معماری سامانه طراحی شده در این پژوهش را تشریح می‌کنیم. در ادامه و در بخش پنج، پس از معرفی مجموعه داده‌های مورد استفاده، خروجی‌های سامانه را بررسی و ارزیابی خواهیم کرد. در انتها و در بخش شش نیز به بررسی نقاط قوت و ضعف سامانه در رنگ‌آمیزی تصاویر و نگاهی به حوزه‌های پژوهشی برای آینده خواهیم پرداخت.

۲- شبکه‌های زیای رقابتی (GAN^۱)

شبکه‌های زیای رقابتی، شبکه‌های عصبی عمیقی هستند که از دو شبکه مجزا و رقیب تشکیل شده‌اند. این شبکه‌ها که نخستین بار در [8] معرفی شدند، توان بالایی در مسائل تولید از یک توزیع داده مشخص دارند. به‌طور ساده می‌توان گفت شبکه‌های رقابتی قادر به زیایی از دامنه‌هایی گسترده همچون تصویر، موسیقی، گفتار یا متن هستند. در این ساختار یک شبکه عصبی عمیق (با نام مؤلفه زیای^۲) به تولید داده‌هایی جعلی می‌پردازد در حالی که شبکه عصبی عمیق دیگری (با نام مؤلفه تمایزی^۳) داده‌های جعلی یادشده را همراه با داده‌های واقعی ارزیابی می‌کند و تصمیم می‌گیرد که آیا داده‌های بررسی شده از مجموعه داده اصلی و واقعی آمده‌اند یا

¹ Generative Adversarial Networks

² Generator

³ Discriminator

جعلی هستند. دامنه عملکرد یک شبکه زیای رقابتی می‌تواند شامل مواردی همچون خلق یک اثر نقاشی، یک موسیقی قابل قبول، یک امضای جعلی، یک صورت انسانی یا طراحی کیف و کفش باشد [9] [10]. موضوع مهم، دسترسی به تعداد قابل قبولی از داده‌های موجود در این دامنه است.

پس از انتخاب یک دامنه مشخص (برای مثال تصاویر ساختمان‌های سنتی یا جلد آلبوم‌های موسیقی) می‌توان مراحل را که یک شبکه زیای رقابتی طی می‌کند، به شرح زیر خلاصه کرد [9]:

۱. مؤلفه زیای، یک عدد تصادفی را به‌عنوان ورودی انتخاب کرده و بر اساس آن یک خروجی جعلی (تصویر) ایجاد می‌کند.

۲. خروجی جعلی مؤلفه زیای به همراه تصویری واقعی از دامنه مورد نظر تحویل مؤلفه تمایزی می‌شود.

۳. مؤلفه تمایزی، تصاویر جعلی و واقعی را گرفته و در انتها احتمالی بین صفر (جعلی بودن تصویر تولیدی) و یک (واقعی بودن تصویر تولیدی) را به دست می‌دهد.

۴. مؤلفه زیای طوری بهینه می‌شود تا مؤلفه تمایزی را فریب دهد (خروجی مؤلفه تمایزی را به سمت یک هدایت کند)

۵. مؤلفه تمایزی نیز طوری بهینه می‌شود که در تشخیص تصاویر جعلی از واقعی استاد شود (خروجی ایش به سمت صفر هدایت شود)

رقابت بین این دو مؤلفه موجب تلاش در حرکت به سمت توزیع واقعی می‌شود؛ اتفاقی که در صورت به‌کارگیری صرف یک مؤلفه زیای به تنهایی رخ نمی‌دهد [8]. در انتهای یک یادگیری موفق، شبکه رقابتی باید قادر به مدل‌سازی مناسبی از یک توزیع تصادفی ورودی به توزیع واقعی دامنه ورودی باشد که این موضوع خود همراه با چالش‌های فراوانی است [11]. نوع خاصی از شبکه‌های زیای رقابتی که بر اساس ورودی مشخصی (همچون یک تصویر به جای یک عدد تصادفی)، خروجی تولید می‌کنند، «شبکه‌های زیای رقابتی شرطی» نام دارند که تفاوتشان با شبکه‌های رقابتی در ورودی آنهاست به طوری که در شبکه‌های شرطی به جای اینکه از اعداد تصادفی برای ورودی استفاده شود، از داده‌ای مشخص و از دامنه‌ای مشخص استفاده می‌شود (همچون یک تصویر یا یک قطعه متن) و هدف، تبدیل داده ورودی به خروجی مورد نظر است. در این ساختار نیز شبکه تمایزی همچون قبل وظیفه تخصیص احتمال به عملکرد شبکه زیای را دارد و آن را

⁴ Conditional GANs

با استخراج اطلاعات از تصویر رنگی و تلفیق آن با تصویر خاکستری، تلاش در رنگ آمیزی تصویر دارد [16]. نقطه ضعف اصلی روش های مبتنی بر راهنمایی کاربر، زمان بر بودن فرایند رنگ آمیزی و پیدا کردن تصویر رنگی مناسب با تصویر خاکستری است. از طرفی مشکل میانگین گیری در روش های مبتنی بر پیدا کردن تصاویر همسان خود را نشان می دهد [17]. با این حال عملکرد سریع تر آن نسبت به رنگ آمیزی دستی موجب شده است که برخی از ابزارهای رنگ آمیزی نوین از این روش ها برای سرعت دهی به رنگ آمیزی بهره گیرند.

ب. رنگ آمیزی با یادگیری^۲: روش های مبتنی بر یادگیری تلاش دارند با بهره گیری از مجموعه داده های تصویری بزرگ، توابعی را برای پیش بینی رنگ پیکسل های درون تصویر فرا گیرند. در این روش ها به مسأله یا به عنوان یک رگرسیون بر روی فضای رنگ پیوسته [7] [18] [20] و یا به عنوان یک مسأله دسته بندی بر روی مقادیر رنگی گسسته نگاه می شود [20].

در پژوهشی دیگر، از دو رویکرد متفاوت و مبتنی بر شبکه های عصبی عمیق برای رنگ آمیزی تصاویر خاکستری بر روی مجموعه داده های Imagenet استفاده شده است [7] [21]. در روش نخست از آپرستون ها^۳ برای به کارگیری تمامی لایه های اطلاعاتی شبکه برای رنگ آمیزی استفاده شده که به خاطر حافظه زیاد مورد نیاز آن، روشی دیگر در این پژوهش به کار گرفته شده است. در روش دوم از تلفیق شبکه VGG^۴ و ساختاری شبیه «شبکه های رسوبی^۵» میکروسافت برای پیش بینی رنگ تصاویر استفاده شده است [22] [23]. با وجود معماری مناسبی که این روش برای رنگ آمیزی ارائه داده، ولی مشکل میانگین گیری در بخش اعظمی از تصاویر رنگ شده قابل رؤیت است. در پژوهشی دیگر تلاش شده است، مشکل میانگین گیری را با تغییر نگرش رگرسیونی برای رنگ آمیزی به نگرشی مبتنی بر دسته بندی رنگ ها و انتخاب رنگ از این فضای دسته بندی حل کنند [6]. در این پژوهش از یک فضای رنگ کوانتیزه شده^۶ استفاده شده و مدل تلاش دارد یکی از ۳۱۳ مقدار قابل انتخاب را از این فضای کوانتیزه برای رنگ آمیزی تصویر انتخاب کند. برای جلوگیری از میانگین گیری و رنگ آمیزی های مُرده، این مدل با یک وزن دهی انطباقی تلاش دارد تا به پیکسل هایی که رنگشان کمتر دیده شده است، وزن بیشتری دهد و در عمل با این کار دسته های رنگی را متوازن کند.

² Learning-based colorization

³ Hypercolumns

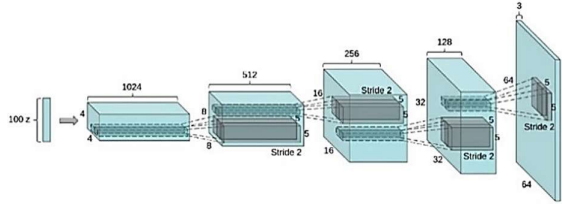
⁴ Visual Geometry Group

⁵ Residual Nets

⁶ LAB

به سمت هدف مورد نظر هدایت می کند [12]. در شکل (۵) نمونه ای از مؤلفه زبای شبکه رقابتی به کار گرفته شده در [9] نمایش داده شده است.

در ادامه و در بخش چهارم به توضیح دقیق تری از چگونگی عملکرد الگوریتم و بهینه سازی ها خواهیم پرداخت.



(شکل-۵): نمونه ای از مؤلفه زبای یک شبکه رقابتی (Figure-5): An example of a generator network in a GAN

۳- پیشینه پژوهش

زمان گیر بودن، عدم مقیاس پذیری و تخصصی بودن رنگ آمیزی به صورت دستی، موجب شکل گیری پژوهش های متعددی برای رنگ آمیزی به کمک مباحث بینایی ماشینی، هوش مصنوعی و یادگیری ماشینی شده است. به رغم از دست رفتن اطلاعات در تصاویر خاکستری، انسان ها قادر به استخراج اطلاعات بافتاری و وابسته به شرایط درون تصویر هستند و بدین ترتیب می توانند تصویر را رنگ آمیزی کنند. این موضوع به این واقعیت اشاره می کند که تصاویر خاکستری دارای اطلاعاتی نهان هستند که می توان از آن ها برای رنگ آمیزی استفاده کرد. در ادامه با نگاهی به پژوهش های انجام شده در حل این مسأله، نقاط قوت و ضعف آن ها را بیان خواهیم کرد.

۳-۱- روش های اصلی در رنگ آمیزی

الف. رنگ آمیزی با راهنمایی کاربر^۱: در پژوهشی با ادغام راهنمایی های کاربر و یک تابع مربع خطا، به رنگ آمیزی تصویر به عنوان یک مسأله بهینه سازی نگاه شده که هدف آن رنگ آمیزی پیکسل های همسایه بر اساس راهنمایی کاربر است [13]. راهنمایی های کاربر محدود به کشیدن خطوط رنگی بر روی تصویر خاکستری است و مدل بدون نیاز به اطلاعاتی دیگر و بر اساس این خطوط و رنگشان، تصویر را رنگ آمیزی می کند. پژوهش های دیگری بر مبنای این روش ارائه شده اند که تلاش در رفع برخی مشکلات همچون سرریز رنگ به قطعات دیگر تصویر و افزایش سرعت آن داشته اند [14] [15]. در رویکردی دیگر، راهنمایی کاربر با ارائه تصویری رنگی و همسان با تصویر خاکستری صورت می گیرد. در این روش مدل

¹ User-guided colorization

با وجود عمق بالا در بیش تر شبکه های یاد شده و به کارگیری معماری های متنوع در طراحی شبکه های یادگیری، در بسیاری از موارد شاهد مشکل میانگین گیری در رنگ آمیزی هستیم. این مشکل بیش از اینکه از ساختار شبکه ها تأثیرپذیر باشد، به دلیل توابع هدفی است که وظیفه هدایت یادگیری شبکه ها را بر عهده دارند. تلاش هایی برای تغییر ساختار توابع هدف انجام شده تا بتوانند سامانه را به استفاده از رنگ های متنوع وادار کند و هرچند توانسته اند به موفقیت نسبی برسند، ولی بحث میانگین گیری و ریزش رنگ به درون دیگر قطعات تصویر هنوز هم در آن ها وجود دارد [6].

۲-۳- میانگین گیری و ضعف توابع خطای

مبتنی بر میانگین مربع خطا (MSE)

رنگ آمیزی تصاویر خاکستری دارای چالش های مهمی همچون «سرریز شدن رنگ»، «ارتباط معنایی اشیا با هم» و «میانگین گیری» است و هدف آن است که سامانه ای طراحی شود که توانایی چیره شدن بر این چالش ها را داشته باشد؛ به همین دلیل نیاز به اشاره و بررسی علت رخداد «میانگین گیری» به عنوان یکی از مهم ترین چالش های مسأله است تا بتوان راه کاری برای خنثی کردن آن ارائه داد. شبکه ای که تلاش در رنگ آمیزی یک تصویر خاکستری دارد و از تابع خطایی همچون MSE برای این کار بهره می برد، چاره ای جز حرکت به سمت میانگین گیری ندارد چرا که این حرکت موجب کاهش خطای کلی شبکه می شود؛ برای توضیح علت این رخداد باید ساختار تابع خطای MSE بررسی شود:

$$L = \frac{1}{2M} \sum_1^T \|\hat{y}_i - y_i\|_2^2 \quad (1)$$

که M تعداد کل تصاویر، T تعداد کل پیکسل های تصویر، \hat{y} تصاویر با رنگ اصلی و y تصاویر رنگ شده است. از طرف دیگر در توزیع گاوسی داریم:

$$p(y|\mu, \sigma^2) = \frac{1}{Z} \exp\left(-\frac{\|y - \mu\|_2^2}{2\sigma^2}\right) \quad (2)$$

در این توزیع Z ثابت هنجارسازی، μ میانگین و σ^2 واریانس است. اگر میانگین توزیع را برابر با داده تصویری ($u = \hat{y}$)، واریانس را واحد ($\sigma^2 = 1$) و ثابت هنجارسازی را نیز نادیده بگیریم (چرا که تأثیر آن در توزیع ها ثابت بوده و می توان آن را حذف کرد) می توان رابطه توزیع را به صورت زیر

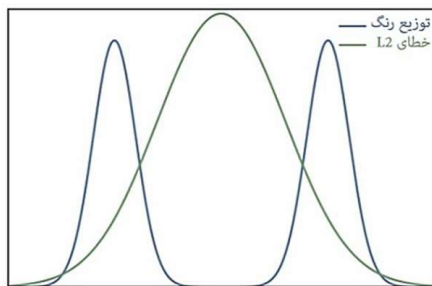
نوشت:

$$p(x|\hat{x}) \propto \exp\left(-\frac{1}{2}\|\hat{x} - x\|_2^2\right) \quad (3)$$

با اعمال لگاریتم، به رابطه زیر می رسیم:

$$\log p(x|\hat{x}) \propto -\frac{1}{2}\|\hat{x} - x\|_2^2 \quad (4)$$

بر اساس روابط (۱ و ۴)، کمینه کردن خطای MSE همچون بیشینه کردن «لگاریتم درست نمایی» یک توزیع گاوسی است با این فرض که داده های تصویری از یک توزیع گاوسی به دست آمده اند. مشکل روش های رنگ آمیزی مبتنی بر تابع خطایی همچون MSE، فرض نادرست تابع بر گاوسی بودن توزیع رنگ است. در بحث رنگ آمیزی تصاویر، با وجود یک تصویر خاکستری اولیه، می توان خروجی های متعددی را به عنوان رنگ آمیزی نهایی تصویر در نظر گرفت، به تعبیر دیگر، تغییر رنگ یک لباس از سبز به آبی همان قدر محتمل و قابل قبول است که تغییر رنگ یک ماشین از قرمز به سیاه. به دلیل وجود چندین حالت مختلف از رنگ آمیزی برای یک تصویر، می توان گفت که توزیع رنگ در تصاویر، توزیعی چندحالتی^۲ است. توزیع گاوسی توزیعی تک حالتی است؛ بدین معنا که در آن فقط یک قله در بین رنگ آمیزی های تصویر وجود دارد و تلاش برای برازش یک توزیع تک حالتی به داده های چندحالتی به کمک تابع خطای MSE، باعث میانگین گیری بین توزیع ها خواهد شد. در شکل (۶) نمونه ای ساده از این اتفاق نشان داده شده است:



(شکل-۶): برازش توزیع تک حالتی به توزیع های چندحالتی (Figure-6): Fitting a unimodal distribution to multi-modal data

اگر در هنگام آموزش، تصاویری را به شبکه دهیم که از توزیع رنگی همانند شکل (۶) گرفته شده باشد، از آنجا که تابع خطای MSE تلاش می کند فاصله رنگ آمیزی مورد نظر را با هر دو نوع رنگ آمیزی موجود کمینه کند، توزیع گاوسی نهایی، توزیعی در میانه توزیع های رنگی واقعی خواهد شد. این موضوع موجب رنگ کردن تصاویر به کمک رنگ هایی می شود

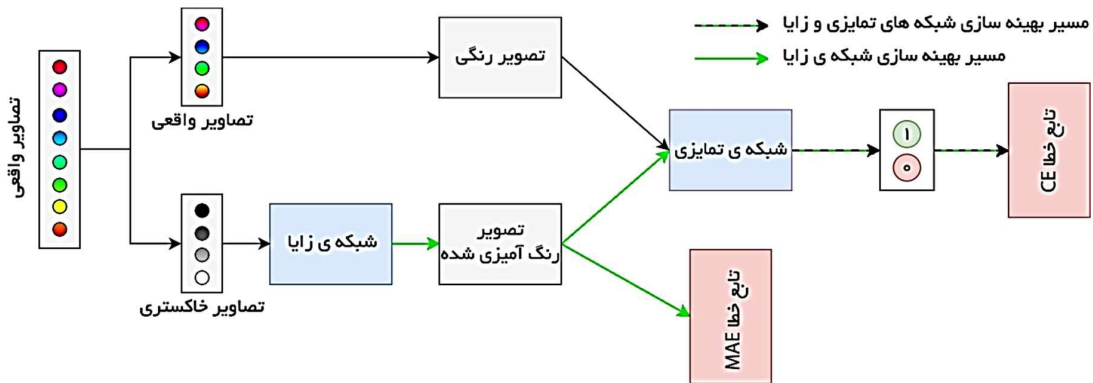
¹ Log-Likelihood

² Multimodal

اصلی (یا تصویری که رنگ‌آمیزی آن واقعی به نظر می‌رسد) داشته باشد. برای این کار، ابزار مورد استفاده، شبکه‌های عصبی عمیق (CNN) و شبکه‌های زایای رقابتی (GAN) است که در ادامه توضیح داده می‌شود.

همچنان که بیان شد، انسان‌ها به‌رغم از دست رفتن اطلاعات در تصاویر خاکستری، قادر به استخراج اطلاعات بافتاری^۱ و وابسته به شرایط درون تصویر هستند تا بر اساس آن بتوانند تصویر را به‌صورتی نسبی رنگ‌آمیزی کنند. استخراج این اطلاعات نهانی از درون تصویر، یکی از اهدافی است که می‌توان برای دست‌یابی به آن از شبکه‌های عمیق بهره برد. بخش اعظمی از موفقیت شبکه‌های CNN به توانایی آن‌ها در یادگیری و تمایز بین رنگ‌ها، الگوها و رابطه‌یابی بین اشکال درون تصاویر برمی‌گردد.

برای غلبه بر مشکل میانگین‌گیری رنگ‌ها، نیاز به تابع خطایی پیچیده‌تر از توابعی همچون MSE است؛ مؤلفه دوم شبکه رقابتی (شبکه تمایزی) که وظیفه تمایز یابی بین تصویر رنگی واقعی و تصویر رنگ‌آمیزی شده را بر عهده دارد، به‌عنوان یک تابع خطای ثانویه برای شبکه زایا عمل می‌کند.



(شکل-۷): ساختار شبکه زایای رقابتی برای رنگ‌آمیزی تصاویر (Figure-7): GAN architecture for colorizing images

است که به هنگام تصویربرداری، از توزیع احتمال خاص آن محیط نمونه‌برداری می‌شود. در این پژوهش تمرکز بر روی رنگ‌آمیزی تصاویری با ابعاد ۲۵۶ در ۲۵۶ است که معادل با تلاش برای نمونه‌برداری از یک توزیع احتمالی ۶۵۵۳۶ بعدی است. این که تصویر «واقعی» است یا نه ارتباط مستقیمی با این توزیع در فضای بالا دارد. مشکل اصلی توزیع تصاویر واقعی است؛ چون به‌غیر از تصاویر نمونه‌برداری شده از توزیع (که زیرمجموعه‌ای کوچک از مجموعه‌ای بسیار بزرگتر و نامتناهی

¹ Generative Adversarial Network
² Contextual Clues

که «واقعی» به نظر نمی‌رسند و در بیش‌تر مواقع حتی بدترین انتخاب ممکن هستند؛ چون که شبکه در عمل تلاش می‌کند رنگ‌هایی را از فضای رنگ‌آمیزی انتخاب کند که نه در توزیع واقعی نخست و نه در توزیع‌های دیگر بلکه در میانه همگی آن‌ها هستند.

این موضوع در رنگ‌آمیزی اشیایی که توزیع رنگ تک‌حالت‌های دارند (همچون چمن یا دریا) تأثیر چندانی ندارد و موجب رنگ‌آمیزی به‌نسبیه موفق این موارد می‌شود (همچنان که در [24] شاهد آن هستیم)، باین‌حال بسیاری از اشیای موجود در دنیای اطراف ما توزیعی چندحالت‌ه برای رنگ خود دارند و همین موجب عملکرد ضعیف بسیاری از سامانه‌های رنگ‌آمیزی در مواجهه با تصاویری همچون ماشین، حیوانات و انسان‌ها می‌شود [7].

۴- مواد و روش‌ها

همچنان که اشاره شد در این مقاله به دنبال آن هستیم که تصویری خاکستری را از یک پرتره انسانی چنان رنگ‌آمیزی کنیم که از دید کاربر نزدیک‌ترین مشابهت را با تصویر رنگی

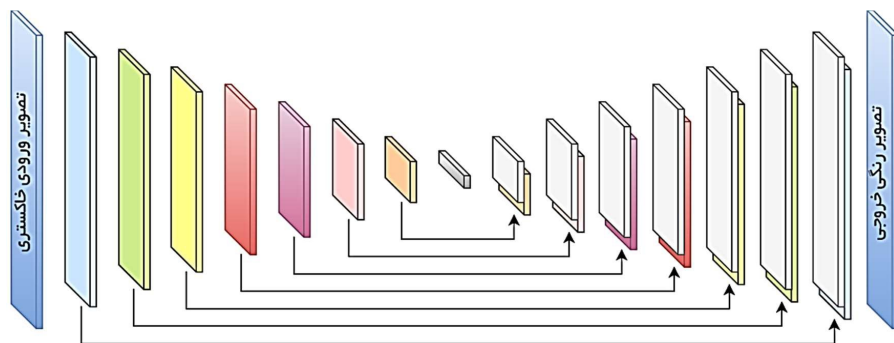
۵- روش پیشنهادی برای رنگ‌آمیزی تصاویر

رنگ‌آمیزی یک تصویر خاکستری را می‌توان تلاشی برای پیدا کردن یک نمونه تصویر رنگ‌شده از فضای تصاویر رنگی یک توزیع احتمالی در نظر گرفت. این موضوع در لحظاتی که در حال خلق یک تصویر هستیم صادق است. به‌عنوان مثال عکاسی از محیط، در عمل به‌عنوان نمونه‌برداری از یک توزیع احتمالاتی پیچیده و در ابعاد بالا (تصویر موردنظر) محسوب می‌شود؛ زیرا هر تصویر شامل تعداد پیکسل‌های مشخصی

از دنیای اطراف است)، هیچ اطلاعاتی در مورد ساختار واقعی آن نداریم.

با توجه به طبیعت مسأله، شبکه موردنظر دارای یک ورودی x (تصویر خاکستری) و خروجی z (تصویر رنگ شده) است. تلاش در پیدا کردن توزیع مناسبی از رنگ‌آمیزی‌های مناسب با توجه به تصویر خاکستری ورودی است که آن را به صورت $p(z|x)$ نمایش می‌دهیم. بدین ترتیب شبکه‌ای رقابتی داریم به طوری که در یک سمت، شبکه زایا تلاش برای رنگ‌آمیزی تصویر خاکستری و فریب دادن شبکه تمایزی و در سمت دیگر شبکه تمایزی تلاش در تشخیص «واقعی» بودن یا نبودن تصویر رنگ‌آمیزی شده دارد. هدف، آموزش شبکه زایا به صورتی است که قادر به رنگ‌آمیزی تصاویری باشد که در نظر شبکه تمایزی واقعی به نظر برسد. در شکل (۷) نمایی از ساختار شبکه رقابتی پیشنهادی نمایش داده شده است. برای ساختار شبکه زایا، از معماری شبیه به آنچه در [7] ارائه شده،

استفاده شده است. این معماری به دلیل ساختار رسوبی خود برتری قابل توجهی نسبت به معماری‌های سنتی ارائه شده بر مبنای شبکه‌های CNN پیشرو^۱ همچون [22] و [25] دارد. معماری‌های سنتی شبکه‌های CNN در وظایف دسته‌بندی و کاهش ابعاد به خوبی عمل می‌کنند [21] ولی در وظایف افزایش ابعاد و ایجاد یک تصویر جدید مشکلات خاص خود را دارند [26]. استفاده از شبکه CNN پیشرو موجب از دست رفتن بسیاری از اطلاعات تصویر به هنگام حرکت در لایه‌های شبکه می‌شود و حضور لایه‌هایی همچون MaxPooling نیز به این مشکل دامن می‌زنند. شبکه استفاده شده در پژوهش از ساختاری رسوبی استفاده می‌کند که نه تنها عملکردی همانند شبکه‌های پیش رو دارد؛ بلکه اضافه شدن لایه‌های رسوبی در هر بخش شبکه موجب جلوگیری از نابودی اطلاعات سطح بالای تصویر و پیشروی این اطلاعات در طول شبکه می‌شود. در شکل (۸) ساختار رسوبی شبکه زایا نشان داده شده است.



(شکل-۸): ساختار رسوبی در شبکه زایا
(Figure-8): Residual architecture in the generator network

زایا را به صورت تلفیق تابع Cross Entropy (مرتبط با شبکه تمایزی) و تابع خطای MAE^2 (بین تصویر رنگ‌آمیزی شده و تصویر واقعی) تعریف می‌کنیم:

$$G_{Loss\ GAN} = -\log(D(\text{Colored Image})) \quad (5)$$

$$G_{Loss\ L1} = \|\text{Real Image} - \text{Colored Image}\|_1 \quad (6)$$

$$G_{Loss\ Total} = 0.5 G_{Loss\ GAN} + 0.5 G_{Loss\ L1} \quad (7)$$

در این روابط Colored Image تصویر رنگ‌آمیزی شده توسط سامانه و Real Image تصویر اصلی دارای رنگ است. علت تلفیق دو تابع خطا، برای بهره‌گیری از هر دو آن‌ها در رنگ‌آمیزی تصاویر است. خطای GAN از رنگ‌آمیزی‌های مرده شبکه پیش‌گیری می‌کند و خطای L1 ساختار کلی

تابع هدف شبکه زایا تلفیقی از دو تابع هدف یکی مبتنی بر خطای بین تصویر رنگ‌آمیزی شده و تصویر واقعی (MAE) و دیگری مبتنی بر خطای بازگشتی از شبکه تمایزی است (Cross Entropy). هم‌چنان که اشاره شد، علت به‌کارگیری تابع MAE به جای MSE در تلاش برای جلوگیری از مات شدن تصاویر ایجاد شده (مشخصه شناخته شده‌ای از MSE) و مردگی بیش از حد رنگ‌آمیزی‌ها است. نخستین لایه شبکه تصویر ورودی هنجار شده بین $[-1, +1]$ را به عنوان یک لایه سه‌بعدی در فضای رنگی LAB می‌گیرد. در هر لایه (به‌غیر از لایه آخر) از تابع فعال‌سازی ReLU و هنجارسازی BatchNorm استفاده شده است. آخرین لایه شبکه برخلاف دیگر لایه‌ها از یک تابع فعال‌سازی تانژانت هایپربولیک (Tanh) استفاده می‌کند تا مقادیر ایجاد شده توسط شبکه زایا را به بازه $[-1, +1]$ محدود کند. از سوی دیگر تابع هدف نهایی شبکه

¹ Feedforward

² Mean Absolute Error

تمایزی به سرعت و به خوبی چگونگی تمایز بین تصاویر ایجاد شده توسط شبکه زایا و تصاویر واقعی را فرامی گیرد. پس از مدتی از شروع کار، شبکه تمایزی در تشخیص تصاویر واقعی از غیر واقعی متخصص شده و مقدار خطای خروجی آن کاهش یافته و گرادیان بازگشتی آن به سمت صفر می‌رود. صفر شدن گرادیان‌ها بدین معناست که شبکه زایا دیگر نمی‌تواند از راهنمایی شبکه تمایزی برای بهبود عملکرد خود بهره‌بردارد و فقط باید بر تابع خطای MAE اتکا کند که در نهایت موجب ایجاد تصاویری با رنگ مرده خواهد شد. برای جلوگیری از این رخداد، ساختار شبکه تمایزی به صورتی ساده و فقط با چندین لایه محدود طراحی شده است. در این مقاله از روش پیشنهادی [8] برای آموزش شبکه زایا (تعداد k برابر بیشتر از شبکه تمایزی) استفاده نشده است؛ زیرا در برخی موارد این موضوع بر پیشرفت یادگیری شبکه اثر منفی دارد.

تصویر را حفظ می‌کند و مانع ریزش رنگ از یک ناحیه به ناحیه نامناسب همسایه می‌شود.

۵-۱- شبکه تمایزی

ساختار شبکه تمایزی بر مبنای یک شبکه CNN سنتی طراحی شده است. یکی از مهم‌ترین مشکلات شبکه‌های رقابتی، واپاشی^۱ شبکه تمایزی است. از یک سو شبکه زایا تلاش می‌کند تصاویر را به صورتی «واقعی» رنگ آمیزی کند و از طرفی شبکه تمایزی سعی در تشخیص تصاویر مصنوعی و غیر واقعی از تصاویر واقعی دارد. مشکل این است که حتی با یک شبکه CNN سنتی و با لایه‌های معدود، وظیفه تمایز بین تصاویر واقعی و تصاویر مصنوعی به راحتی قابل انجام است. از طرفی شبکه زایا به سختی می‌تواند رنگ آمیزی تصاویر را بر مبنای گرادیان بازگشتی بهینه کند و در طرف دیگر شبکه

for number of iterations do

- Sample m real color images $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ from real color distribution $p_{data}(x)$
- Generate m fake colored images $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ from real gray distribution $p_{data}(x)$
- Update the discriminator:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log(D(x^{(i)})) + \log(1 - D(G(z^{(i)})))]$$

- Generate m fake colored images $\{z^{(1)}, z^{(2)}, \dots, z^{(m)}\}$ from real distribution $p_{data}(x)$
- Update the generator:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m [\log(1 - D(G(z^{(i)})))]$$

(شکل-۹): شبه‌کد یادگیری دو شبکه
(Figure-9): Pseudo-code of the two networks

شبکه زایا را یاد می‌گیرد. در اینجا هدف شبکه تمایزی دو نکته زیر است:

۱. بیشینه کردن $D(x)$ برای هر تصویری که از توزیع رنگی واقعی $x \sim p_{data}$ است.
۲. کمینه کردن $D(x)$ برای هر تصویری که از توزیع رنگی واقعی $x \sim p_{data}$ نیست.

این دو هدف را می‌توان به صورت خلاصه زیر در نظر گرفت:

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}} \log[D(x)] + \mathbb{E}_{z \sim p_z} \log[1 - D(G(z))] \quad (11)$$

شبه‌کد نهایی برای یادگیری دو شبکه به صورت شکل (۹) خواهد بود. اطلاعات بیشتر در مورد جزئیات مدل و شمایی از ساختار کلی آن در شکل (۱۰) و جداول (۱) و (۲) آمده است. ورودی شبکه زایا، تصویر خاکستری (تک‌کاناله) و خروجی شبکه زایا تصویر رنگ آمیزی شده جعلی (تصویر سه کاناله)

۵-۲- تابع هدف

تابع هدف نهایی شبکه تمایزی را می‌توان به صورت تلفیق تابع Cross Entropy برای تصاویر واقعی و غیر واقعی نوشت:

$$D_{Loss Real} = -\log(y | \text{Real Image}) \quad (8)$$

$$D_{Loss Fake} = -\log(1 - y | \text{Colored Image}) \quad (9)$$

$$D_{Loss Total} = 0.5 D_{Loss Real} + 0.5 D_{Loss Fake} \quad (10)$$

y مقدار پیش‌بینی شبکه تمایزی (عددی بین صفر و یک) در مورد واقعی بودن یا نبودن تصویر رنگ آمیزی شده است.

۵-۳- الگوریتم یادگیری

ایده اصلی یادگیری شبکه‌های رقابتی بر مبنای بازی MinMax است؛ به طوری که در وهله نخست شبکه تمایزی مفاهیم بین تصویر واقعی و تصویر رنگ آمیزی شده (توسط

^۱ Collapse

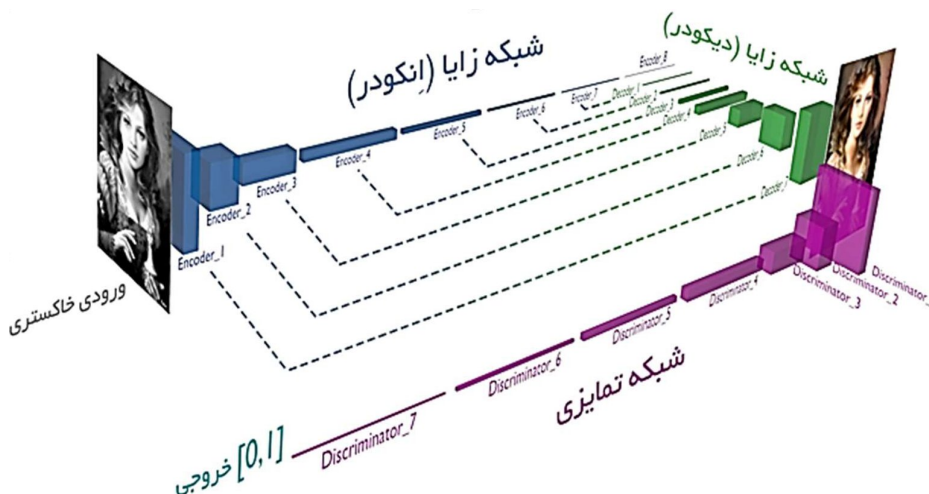
به کارگیری چندین مجموعه داده متفاوت، تلاش برای پوشش دامنه‌های تصویری متفاوت است. به عنوان مثال تصاویر مجموعه دویست هزار تایی CelebA تصاویری اغلب بدون نوفه و با چارچوب به طور تقریبی یکسان هستند؛ به طوری که تصویر افراد در این مجموعه اغلب میانه بالایی بدن است و تصاویر پس زمینه، کمترین بخش تصاویر را در این مجموعه تشکیل می‌دهد. ساختار این مجموعه آن را برای رنگ آمیزی صورت‌ها و لباس‌های افراد از نزدیک مناسب کرده است؛ در حالی که اگر تصویر از فاصله‌ای دورتر گرفته شده باشد، این مجموعه چندان به یادگیری رنگ آمیزی افراد و پس زمینه تصاویر کمک نمی‌کند.

است. ورودی شبکه تمایزی نیز تلفیق هر دو تصویر با رنگ آمیزی واقعی و جعلی است (دو تصویر سه کاناله). کل پیکسل‌های تصویر بصورت یک ماتریس $1 \times 256 \times 256$ به شبکه اعمال شده است و به خاطر نیاز به شناسایی روابط کلی بین اشیای درون تصویر از اعمال ورودی‌ها به صورت تکه تکه به شبکه پرهیز شده است. به خاطر نیاز به کانون‌لشون بر روی تصاویر دوبعدی از اعمال آنها به شبکه به صورت برداری نیز پرهیز می‌شود.

۶- پیاده‌سازی و ارزیابی عملکرد سامانه

۶-۱- داده‌ها

برای آموزش شبکه، تلفیقی از مجموعه داده‌های CelebA [27] و MS-Celeb [28] میکروسافت استفاده شده است. علت



(شکل-۱۰): ساختار شبکه

(Figure-10): Network Architecture

(جدول-۱): جزئیات شبکه زایا

(Table-1): Generator network details

نام لایه	اندازه تصویر	تعداد کانال	نوع کانال	توضیحات
تصویر خاکستری	256×256	1		تصویر خاکستری کانال L از $LA * B * L$
Encoder_1	128×128	64	Conv + LReLU	
Encoder_2	64×64	128	Conv + BatchNorm + LReLU	
Encoder_3	32×32	256	Conv + BatchNorm + LReLU	
Encoder_4	16×16	512	Conv + BatchNorm + LReLU	
Encoder_5	8×8	512	Conv + BatchNorm + LReLU	
Encoder_6	4×4	512	Conv + BatchNorm + LReLU	
Encoder_7	2×2	512	Conv + BatchNorm + LReLU	
Encoder_8	1×1	512	Conv + BatchNorm + LReLU	
Decoder_1	2×2	512 + 512	DeConv + BatchNorm + ReLU	تلفیق با Encoder 7
Decoder_2	4×4	512 + 512	DeConv + BatchNorm + ReLU	تلفیق با Encoder 6
Decoder_3	8×8	512 + 512	DeConv + BatchNorm + ReLU	تلفیق با Encoder 5
Decoder_4	16×16	512 + 512	DeConv + BatchNorm + ReLU	تلفیق با Encoder 4

Decoder_5	32 × 32	256 + 256	DeConv + BatchNorm + ReLU	تلفیق با Encoder 3
Decoder_6	64 × 64	128 + 128	DeConv + BatchNorm + ReLU	تلفیق با Encoder 2
Decoder_7	128 × 128	64 + 64	DeConv + BatchNorm + ReLU	تلفیق با Encoder 1
Decoder_8	256 × 256	2	DeConv + Tanh	کانالهای A* و B*
تصویر رنگ شده	256 × 256	3	3	تلفیق کانالهای L, A*, B*

(جدول ۲-): جزئیات شبکه تمایزی
(Table-2): Discriminator network details

نام لایه	اندازه تصویر	تعداد کانال	نوع کانال	توضیحات
تصویر رنگ شده + تصویر واقعی	256 × 256	3 + 3		تلفیق دو تصویر رنگ شده و واقعی
Discriminator_1	128 × 128	64	Conv + LReLU	
Discriminator_2	64 × 64	128	Conv + BatchNorm + LReLU	
Discriminator_3	32 × 32	256	Conv + BatchNorm + LReLU	
Discriminator_4	16 × 16	512	Conv + BatchNorm + LReLU	
Discriminator_5	8 × 8	512	Conv + BatchNorm + LReLU	
Discriminator_6	4 × 4	512	Conv + BatchNorm + LReLU	
Discriminator_7	2 × 2	512	Conv + BatchNorm + LReLU	
مقدار خروجی	1 × 1	1	Conv + Sigmoid	خروجی بین ۰ و ۱

تصویر خاکستری



Zhang et al (2016)



روش پیشنهادی



(شکل ۱۱): رنگ آمیزی تصاویری که رنگ واقعی آنها موجود نیست
(Figure-11): Colorized images without any "real" color available

لحاظ رنگ آمیزی و هم از لحاظ به کارگیری در ساختار شبکه دارد [24]. علاوه بر این کانال L در این فضا در عمل تصویر خاکستری (میزان شدت نور) مورد نظر ماست که می توان آن را به عنوان ورودی شبکه در نظر گرفت و دو کانال A و B نیز که رنگ بندی تصویر را رقم می زنند، خروجی مورد نظرمان از شبکه زایا هستند. با این کار وظیفه شبکه را برای پیش بینی سه کانال رنگ (قرمز، سبز و آبی در فضای RGB) به پیش بینی دو کانال (A و B) تقلیل داده ایم.

از طرفی مجموعه یک میلیونی MS-Celeb شامل تصاویر افراد در دامنه های متفاوت به همراه تصاویری است که در عمل نوفه محسوب می شوند و البته در برخی از آن ها افراد نیز حضور ندارند. این مجموعه کمک می کند پس زمینه تصاویر بهتر رنگ آمیزی شود. برای ایجاد توازن بین دو مجموعه، دو بیست هزار تصویر از CelebA و دو بیست هزار تصویر دیگر از MS-Celeb انتخاب شده و اندازه آن ها به ابعاد ۲۵۶ در ۲۵۶ کاهش یافته است. فضای رنگی تصاویر را از RGB به فضای LAB تغییر می دهیم؛ چون که این فضا مزایایی جدی، هم از

۲-۶- آزمایش تجربی

از مدتی موجب واپاشی شبکه تمایزی می‌شود. در این بین روش Adam با آهنگ یادگیری استاندارد 0.0001 و مقدار پتانسی 0.16 خروجی مناسبی را در بین آزمایش‌ها نشان می‌دهد. در شکل (۱۱) شاهد رنگ‌آمیزی برای تصاویری هستیم که هیچ رنگ‌آمیزی واقعی از آن‌ها در دسترس نیست. در تصویر ستون نخست (از سمت چپ)، با نگاهی به رنگ پیراهن لورل و هاردی می‌توان رخداد میانگین‌گیری و رنگ سوبیا را در جایی که پیراهن باید سفید باشد، تشخیص داد، این مشکل در حالی در روش Zhang وجود دارد. ولی در رویکرد پیشنهادی رخ نمی‌دهد.

شبکه رقابتی را با تنظیمات متفاوتی برای پنج دور کامل و با دسته‌های ده‌تصویره ($Batchsize=10$) آزمایش کردیم. حافظه محدود و سرعت به‌نسبیه پایین کلاک پردازنده موجب شد که هر آزمایش به‌طور تقریبی یک هفته زمان ببرد. برای الگوریتم یادگیری، روش‌های Adam [30]، AdaDelta [29] و SGD بررسی شدند. تنظیم دستی و مناسب آهنگ یادگیری برای SGD فرایندی زمان‌بر و همراه با آزمون و خطاست و از طرفی باعث واپاشی شبکه تمایزی می‌شود و یا میزان یادگیری آن بسیار کند است. در مقابل روش AdaDelta پس



(شکل-۱۲): تصاویر رنگ‌آمیزی شده به همراه رنگ واقعی
(Figure-12): Colorized images available color

درخت شناسایی کند؛ ولی به‌اشتباه بدنه درخت را نیز همانند برگ‌ها به رنگ سبز رنگ‌آمیزی کرده است. از طرفی رنگ‌آمیزی فرد داخل تصویر به‌صورت سوبیا و با میانگین‌گیری انجام شده، در حالی که در این روش نه تنها پوست فرد درست رنگ شده، بلکه دامن وی نیز دارای رنگی به‌طور کامل متفاوت از سوبیا شده است و همچنین رنگ بدنه درخت با رنگ سبز برگ‌های پشت آن مخلوط نشده است. مشکل میانگین‌گیری در ستون آخر نیز مشهود است؛ در حالی که در روش پیشنهادی تلفیقی از رنگ‌های قرمز، سبز و آبی برای رنگ‌آمیزی تصویر به کار رفته‌اند و شبکه بر جزئیات ریزی همچون چشم فرد (سبز رنگ) نیز دقت یافته است. در شکل (۱۲) شاهد

در تصویر ستون دوم، شبکه به‌هیچ‌وجه نتوانسته عناصر درون تصویر و رنگ‌آمیزی مناسب آن‌ها را تشخیص دهد و با پاشیدن رنگ سوبیا بر روی کل آن تلاش کرده تا میزان خطای کل را کاهش دهد؛ در حالی که شبکه پیشنهادی نه تنها فرد را رنگ‌آمیزی کرده، بلکه حتی کتاب‌های کتابخانه را نیز از قلم نینداخته است. در تصویر ستون سوم نیز مشکل میانگین‌گیری خودنمایی می‌کند؛ با این تفاوت که ماشین به‌جای رنگ به‌طور کامل سوبیا، رنگی بین سوبیا و قرمز گرفته است در حالی که در روش پیشنهادی این مشکل وجود ندارد و از فضای رنگ و به‌خصوص رنگ قرمز به‌خوبی استفاده شده است. در ستون چهارم، مدل Zhang توانسته برگ‌ها را در پشت

بوجود عملکرد مناسب شبکه در بسیاری از تصاویر که ساختاری استاندارد دارند (مانند صورت، اندازه فرد، کیفیت تصویر) باین حال مشکلاتی در رنگ‌آمیزی بازه متفاوتی از تصاویر نیز وجود دارد. به‌عنوان مثال در ستون نخست (از سمت چپ)، لباس زن توسط شبکه، صورتی رنگ‌آمیزی شده در حالی که با دقت در جزئیات لباس متوجه می‌شویم که فقط نیمه راست آن رنگ‌آمیزی شده و نیمه چپ لباس با رنگی تلفیقی از سوبیا و سبز پوشانده شده است. یا به‌عنوان مثالی دیگر در تصویر ستون آخر قسمتی از دست ورزشکار همانند لباس وی رنگ‌آمیزی شده است.

رنگ‌آمیزی برای تصاویری هستیم که نسخه واقعی آن‌ها موجود است.

همچنان که ملاحظه می‌شود، برخلاف روش Zhang که به‌طور تقریبی در تمامی حالات از پاشیدن رنگ سوبیا برای رنگ‌آمیزی بهره برده، روش پیشنهادی نه تنها از رنگ‌های زنده و غیر سوبیا برای رنگ‌آمیزی استفاده کرده بلکه به جزئیاتی همچون رنگ چشم و لب‌ها نیز توجه شایانی شده است تا جایی که به‌عنوان مثال در تصویر ستون سوم، رنگ‌آمیزی ارائه شده توسط شبکه بهتر و زنده تر از رنگ‌آمیزی دنیای واقعی به نظر می‌رسد.



(شکل-۱۳): تصاویر رنگ‌آمیزی شده با مشکلات مختلف
(Figure-13): Common problems in colorization

به کمک Adobe Photoshop رنگ شده است) باین‌حال احتمال رنگ‌آمیزی به شکلی که شبکه انجام داده به شدت پایین است. ستون سوم رنگ‌آمیزی ناپیوسته دیگری را نشان می‌دهد که برخلاف ستون دوم ناپیوستگی معنایی و سطح بالا دارد تا ناپیوستگی فیزیکی. در این تصویر قسمتی از بدن به همراه دست فرد به درستی رنگ‌آمیزی شده در حالی که دست دیگر وی به رنگ سبز و به‌طور کامل متفاوت از رنگ واقعی آن رنگ‌آمیزی شده است. در واقع شبکه در این تصویر بیش از آنکه به ساختار معنایی قطعات جداگانه (از جمله دست‌ها) اهمیت دهد، تلاش داشته تا بر اساس میزان شدت روشنایی، رنگ‌آمیزی کند و از آنجاکه دست راست فرد شدت رنگی متفاوت با صورت و دست چپش دارد، رنگ‌آمیزی متفاوتی برای آن انجام شده است. ستون چهارم تلفیقی از مشکلات یادشده در تصاویر قبل را به همراه دارد که در بین آن‌ها می‌توان به رنگ‌آمیزی ناپیوسته فیزیکی در پس‌زمینه و

در شکل (۱۳) نمونه‌هایی منتخب از عملکرد ضعیف شبکه در رنگ‌آمیزی نمایش داده شده است، بسیاری از مشکلات در رنگ‌آمیزی تصاویر، قابل خلاصه‌شدن در مشکلات یادشده در این شکل است.

در ستون نخست (از سمت چپ)، شاهد ورزشکاری هستیم که مدال خود را بالا آورده است. ناحیه کوچکی از گردن وی بین لباس و بند مدالش محصور شده که شبکه با اشتباه در شناسایی این قسمت از گردن او، این ناحیه را به رنگی همانند رنگ لباسی که برایش انتخاب کرده، رنگ‌آمیزی کرده است. این مشکل حکایت از ضعف شبکه در شناسایی کامل ساختار بدن انسان و چگونگی پوشش آن دارد. در ستون دوم رنگ‌آمیزی لباس به‌صورتی ناپیوسته انجام شده به‌طوری که قسمتی از آن به رنگ قرمز و قسمتی دیگر به رنگ آبی در آمده است، با اینکه تصویر رنگی این ستون، تصویری واقعی نیست (در گروه r/Colorization توسط متخصص و

قطعات کلیدی تأثیرگذاری همچون انباشته شدن برگ درختان بر روی زمین (و نه بر روی شاخه‌ها) و بی‌برگ بودن درختان، نشان از آب‌وهوای پاییزی تصویر دارد. در چنین مواقعی شبکه بیش از اینکه به این اشارات و روابط بین آن‌ها اهمیتی بدهد، تمرکز را بر شناسایی اجزای تصویر قرار می‌دهد که در نهایت باعث می‌شود، شاهد رنگ‌آمیزی سبز و زنده‌ای برای برگ‌های مُرده انباشته‌شده بر روی چمن‌های با طراوت باشیم. یا در نمونه‌ای دیگر ممکن است، تصویری خاکستری از افراد با لباس فرم هم‌رنگ داشته باشیم که به دلیل نورپردازی محیط، هر لباس شدت روشنایی متفاوتی گرفته باشد. در این سناریو شبکه بیش از اینکه به ساختار یکسان لباس‌های فرم اهمیت بدهد، بر اساس شدت روشنایی متفاوتشان، رنگ‌های متنوعی را به آن‌ها اختصاص می‌دهد و این موضوع باعث می‌شود که در تصویری از افراد با لباس‌های یکسان، شاهد رنگ‌های متفاوتی از آبی و قرمز تا سبز و صورتی باشیم.

همچنین رنگ‌آمیزی ناپیوسته مفهومی برای دست فرد (به رنگ آبی) اشاره کرد. ستون پنجم نمایش‌دهنده طیفی از تصاویر است که شبکه، آن‌ها را به صورت سوبیا رنگ‌آمیزی می‌کند.

در این تصاویر شبکه نتوانسته اجزای تأثیرگذار در رنگ‌آمیزی (از جمله آسمان، چمن و یا افراد) را شناسایی کند. پیچیده بودن ساختار تصویر، فاصله زیاد دوربین از افراد و کوچک شدن اشیای قابل‌شناسایی در تصویر از جمله عوامل تأثیرگذار در ناکارآمدی شبکه در رنگ‌آمیزی چنین تصاویری هستند. در چنین سناریوهایی بخش زیادی از شبکه تصویر را سوبیا رنگ کرده و رنگی مُرده را به ارمغان می‌آورد. در شکل (۱۴) نمونه‌هایی از این عملکرد ناقص نمایش داده شده است. به غیر از مشکلاتی که در بالا بدان‌ها اشاره کردیم، در برخی مواقع شاهد نمونه‌هایی از خطا در رنگ‌آمیزی هستیم که نشان از درک مفهومی ضعیف شبکه از معانی دنیای اطراف ما دارد. به عنوان مثال در تصویری خاکستری از یک پارک،

تصویر خاکستری



روش پیشنهادی

(شکل-۱۴): تصاویر با رنگ‌آمیزی کاملاً نامناسب

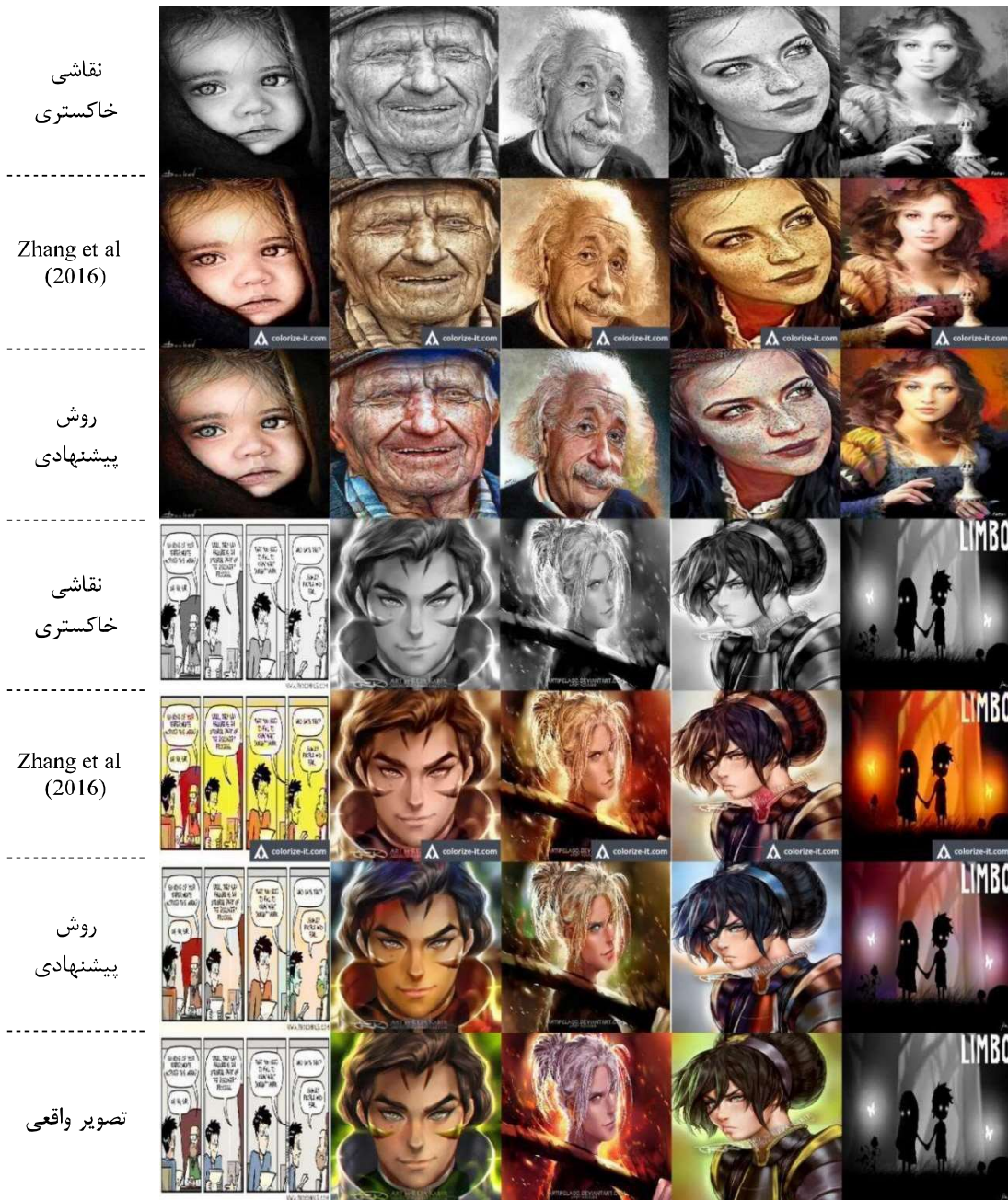
(Figure-14): Failed colorizations

در ردیف نخست شاهد نقاشی‌هایی هستیم که با دست کشیده شده‌اند. در بسیاری از موارد طرح‌های هنری به عمد رنگ‌آمیزی نمی‌شوند تا ساختار آن واقعی‌تر به نظر برسد؛ چون رنگ‌آمیزی آن‌ها ممکن است از رئالیسم موجود در طرح بکاهد. روش پیشنهادی با این موضوع مشکلی ندارد و رنگ‌آمیزی بسیار زنده‌ای برای این تصاویر ارائه می‌دهد تا جایی که بدون اطلاع از خاستگاه تصاویر، حتی ممکن است آن‌ها را با تصاویری واقعی اشتباه گرفت. در ردیف سوم به پایین تصاویری با تنوع بیشتر بررسی شده‌اند. در ستون نخست تصویر یک وب‌کامیک به نمایش درآمده است.

برای بررسی انعطاف نسبت به تصاویر غیرانسانی از بازه‌ای از نقاشی‌ها، کامیک‌ها و بازی‌های ویدیویی بهره گرفتیم. در حوزه بازی‌های ویدیویی بسیاری از نویسه‌ها به خوبی رنگ‌آمیزی می‌شوند؛ چون طراحی صورت و بدن بیش‌تر آن‌ها همانند انسان‌های واقعی است؛ ولی در حوزه تصاویر هنری این هارمونی وجود ندارد که باعث می‌شود، شبکه در برخی از آن‌ها قادر به رنگ‌آمیزی موفق باشد و در برخی دیگر خروجی‌های نه‌چندان مناسبی بدهد. در شکل (۱۵) برخی از رنگ‌آمیزی‌های موفق در تصاویر هنری و نقاشی نشان داده شده‌اند.

است که از لحاظ کیفی با آنچه نقاش ساعت‌ها برای آن زمان صرف کرده، برابری می‌کند. تصویر ستون آخر از بازی ویدئویی Limbo گرفته شده که محیطی به‌طور کامل خاکستری داشته و در هیچ جای بازی رنگ به کار نرفته است؛ ولی شبکه توانسته تصویر را، هر چند به‌شکل محدود، رنگ‌آمیزی کند.

بسیاری از هنرمندان زمان کافی را هم برای کشیدن و هم برای رنگ‌آمیزی یک کامیک ندارند و می‌توان از قابلیت‌های چنین مدلی برای رنگ‌آمیزی تصاویر خاکستری آن‌ها بهره برد. ستون‌های دوم تا چهارم تصاویری از یک هنرمند را نمایش می‌دهند. همان‌طور که ملاحظه می‌شود شبکه رنگ‌آمیزی جدیدی را برای هر سه تصویر ارائه داده



(شکل-۱۵): رنگ‌آمیزی تصاویر هنری

(Figure-15): Colorizing artistic images

رنگ‌آمیزی را به کمک توابع خطای ساده‌ای همچون MSE (و یا دیگر خطاهای مطرح) سنجید. به‌عنوان مثال در مسئله تشخیص عمق، سنجش میزان خطای عمق خروجی با عمق

در حوزه بررسی کارایی با مسأله‌ای حل‌نشده روبه‌رو هستیم. برخلاف مسائلی همچون کاهش نوفه، افزایش ابعاد و یا تشخیص عمق در تصاویر، نمی‌توان کارایی یک مسئله

واقعی به‌تنهایی و به‌کمک MSE قابل‌محاسبه است؛ چون خروجی‌های مورد‌انتظارمان در چنین مسائلی فقط یک تصویر در دو بُعد و بدون هیچ‌گونه انتزاع معنایی هستند؛ درحالی‌که در بحث رنگ‌آمیزی، هدف رسیدن به رنگ‌بندی واقعی تصویر نیست، بلکه هدف رنگ‌آمیزی به‌صورتی است که تصویر را «واقعی» جلوه دهد و به‌همین دلیل با به‌کارگیری توابع خطای معمول (همچون MSE) ممکن است، حتی رنگ‌آمیزی «واقعی» نیز دچار خطای بالایی شود. به‌عنوان مثال تغییر رنگ یک پیراهن آبی‌رنگ (رنگ واقعی) با قرمز موجب افزایش خطا خواهد شد، حتی اگر پیراهن قرمز، «واقعی» نیز به نظر برسد. با توجه به مسأله بالا، تابع خطایی لازم است که قادر به تشخیص میزان «واقعیت و کیفیت» رنگ‌آمیزی باشد. نزدیک‌ترین ایده‌ای که برای این کار به ذهن خطور می‌کند، استفاده از یک شبکه (به‌جای یک تابع خطای ساده) برای تشخیص موضوع است که همان کاری را انجام می‌دهد که شبکه تمایزی پیشنهادی انجام می‌دهد.

۷- نتیجه‌گیری

در این مقاله سامانه‌ای هوشمند برای رنگ‌آمیزی تصاویر خاکستری در دامنه پرتره افراد معرفی شد. سامانه پیشنهادی به‌صورت «پشت سر هم» طراحی شده است که تصویری خاکستری را از ورودی گرفته و رنگ‌آمیزی متناسبی برای آن ارائه می‌کند. این سامانه از دو شبکه متصل به هم تشکیل شده است که شبکه نخست (زایا) وظیفه ارائه یک رنگ‌آمیزی منتخب برای تصویر و شبکه دوم (تمایزی) وظیفه برآورد و سنجش عملکرد شبکه نخست و راهنمایی آن را دارد. در طول روند یادگیری و با شکل‌گرفتن یک بازی MinMax در بین شبکه‌ها، عملکرد هر دوی آنها رو به بهبودی می‌رود؛ به‌طوری‌که شبکه زایا رنگ‌آمیزی‌های جعلی بهتری می‌کند و شبکه تمایزی نیز بین رنگ‌آمیزی‌های واقعی و غیر واقعی تمایز دقیق‌تری قائل می‌شود. سامانه بر روی چهارصد هزار تصویر متشکل از دو مجموعه داده آموزش‌یافته و عملکردش بر روی تصاویری متفاوت از دامنه‌های تصاویر آموزشی، بررسی شده است. در چالش میانگین‌گیری، عملکرد سامانه نسبت به سامانه‌های دیگر بررسی شد و برتری آن در دوری کردن از توزیع رنگ‌های مرده نشان داده شد. از طرفی سامانه قابلیت رنگ‌آمیزی آثار هنری در حوزه تصاویر پرتره را نیز به‌خوبی از خود نشان داده است. با این‌حال در مواردی که نیاز به درک

معنایی از محتوای درون تصویر است، سامانه به‌خوبی عمل نمی‌کند و مسأله درک معنایی تصاویر در سطح انسان و رنگ‌آمیزی بر مبنای آن هنوز مسأله‌ای باز در این حوزه محسوب می‌شود؛ از این‌رو برای پژوهش‌های آینده می‌توان به موارد زیر اشاره کرد:

بخش اعظمی از عملکرد شبکه تحت تأثیر زمان و حافظه محدودی است که یادگیری در آن انجام شده است؛ از این جهت پیشنهاد می‌شود برای افزایش کارایی آن از کل مجموعه داده‌های بزرگی همچون MS Celeb و در زمانی معقول به همراه دسته‌های تصویری بزرگ (تا جایی که حافظه اجازه می‌دهد) استفاده شود. با توجه به ساختار فضای رنگی LAB، می‌توان سامانه را محدود به تصاویر ۲۵۶ در ۲۵۶ نکرد و ابتدا رنگ‌آمیزی لایه‌های AB را استخراج کرد و سپس به‌کمک روش‌های افزایش ابعادی همچون «درون‌یابی دوخطی»^۲ این فضای رنگی را به‌اندازه واقعی تصویر تغییر بُعد داده و در نهایت آن را بر روی لایه L سوار کرد، بدین ترتیب رنگ‌آمیزی را بر روی تصاویری با ابعاد بزرگ‌تر از ۲۵۶ نیز می‌توان انجام داد. با وجود بررسی و آزمایش تابع خطای Wasserstein [11] برای شبکه تمایزی، به خروجی قابل‌قبولی بر مبنای این توابع نرسیدیم و شبکه در نهایت به میانگین‌گیری از رنگ‌ها و پاشی می‌کرد؛ ولی با در نظر گرفتن برتری این تابع خطا نسبت به توابع خطای مبتنی بر دیورژانس Kullback-Leibler در جلوگیری از واپاشی شبکه تمایزی، پژوهش‌های بیشتر در این حوزه باید انجام شود.

8- References

۸- مراجع

- [1] J. J. Lloyd, "The Complexity of Recolouring Photos," 2017. [Online]. Available: <https://www.fxguide.com/featured/the-complexity-of-re-colouring-photos/>.
- [2] "r/colorizationrequests." [Online]. Available: <https://www.reddit.com/r/colorizationrequests>.
- [3] P. Whitt, *Pro Photo Colorizing with GIMP*. Apress, 2016.
- [4] S. Koo, "Automatic Colorization with Deep Convolutional Generative Adversarial Networks," 2016. [Online]. Available: http://cs231n.stanford.edu/reports2016/224_Report.pdf.
- [5] Aleju, "Aleju Torch Colorizer," 2016. [Online]. Available: <https://github.com/alcju/colorizer>.
- [6] R. Zhang, P. Isola, and A. A. Efros, "Colorful Image Colorization," *Eccv*, pp. 1–25, 2016.
- [7] R. Dahl, "Automatic Colorization," 2016.

² Bilinear Interpolation

¹ End to End

- [22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Iclr*, vol. 96, no. 2, pp. 1–14, 2015.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Arxiv.Org*, vol. 7, no. 3, pp. 171–180, 2015.
- [24] S. Iizuka, Edgar Simo-Serra, and H. Ishikawa, "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification," *Siggraph '16*, vol. 35, no. 4, pp. 1–11, 2016.
- [25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Nips*, 2012, pp. 1–9.
- [26] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and Checkerboard Artifacts," *Drill*, pp. 1–14, 2016.
- [27] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 11–18–Dece, pp. 3730–3738, 2016.
- [28] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World," *Eur. Conf. Comput. Vis.*, pp. 87–102, 2016.
- [29] M. D. Zeiler, "ADADELTA: An Adaptive Learning Rate Method," *arXiv*, p. 6, 2012.
- [30] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *Int. Conf. Learn. Represent.*, 2014.
- [Online]. Available: <http://tinyclouds.org/colorize/>.
- [8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," Jun. 2014.
- [9] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv*, pp. 1–15, 2015.
- [10] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative Adversarial Text to Image Synthesis," *Icml*, pp. 1060–1069, 2016.
- [11] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," Jan. 2017.
- [12] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *CoRR*, pp. 1–7, 2014.
- [13] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Trans. Graph.*, vol. 23, no. 3, p. 689, 2004.
- [14] Y.-C. Huang, Y.-S. Tung, J.-C. Chen, S.-W. Wang, and J.-L. Wu, "An adaptive edge detection based colorization algorithm and its applications," *Proc. 13th Annu. ACM Int. Conf. Multimed. - Multimed. '05*, no. January, p. 351, 2005.
- [15] L. Yatziv and G. Sapiro, "Fast image and video colorization using chrominance blending," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1120–1129, 2006.
- [16] T. Welsh, M. Ashikhmin, and K. Mueller, "Transferring color to greyscale images," *ACM Trans. Graph.*, vol. 21, no. 3, pp. 277–280, 2002.
- [17] R. Gupta, A. Chia, and D. Rajan, "Image colorization using similar images," *Proc. 20th ...*, pp. 369–378, 2012.
- [18] Z. Cheng, Q. Yang, and B. Sheng, "Deep colorization," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 11–18–Dece, pp. 415–423, 2016.
- [19] A. Deshpande, J. Rock, and D. Forsyth, "Learning large-scale automatic image colorization," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 11–18–Dece, pp. 567–575, 2016.
- [20] G. Charpiat, M. Hofmann, and B. Schölkopf, "Automatic image colorization via multimodal predictions," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5304 LNCS, no. PART 3, pp. 126–139, 2008.
- [21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.



محمد مهدی حاجی اسمعیلی، مدرک

کارشناسی فناوری اطلاعات خود را در سال ۱۳۹۱ از دانشگاه صنعتی شاهرود و مدرک کارشناسی ارشد خود را نیز در رشته یادشده و در گرایش تجارت

الکترونیک در سال ۱۳۹۳ از دانشگاه قم دریافت کرد. موضوع پایان نامه کارشناسی ارشد ایشان طراحی و پیاده سازی سامانه ای برای ارتباط با پایگاه های داده تجاری به واسطه زبان طبیعی بوده است. وی هم اکنون دانشجوی مقطع دکترا در همان رشته و در دانشگاه تربیت مدرس است. زمینه های پژوهشی مورد علاقه وی یادگیری ماشینی، یادگیری تقویتی، پردازش تصویر و پردازش زبان طبیعی است.

نشانی رایانه ایشان عبارت است از:

MohammadHaji@Modares.ac.ir



غلامعلی منتظر، مدرک کارشناسی

مهندسی برق خود را در سال ۱۳۷۰ از

دانشگاه خواجه نصیرالدین طوسی

دریافت کرد. وی مدارک کارشناسی

ارشد و دکترای خود را نیز در رشته یاد

شده در سال‌های ۱۳۷۳ و ۱۳۷۷ از دانشگاه تربیت مدرس

دریافت کرد. ایشان هم‌اکنون دانشیار گروه مهندسی فناوری

اطلاعات دانشگاه تربیت مدرس است. زمینه‌های پژوهشی

مورد علاقه وی هوش مصنوعی، روش‌های مبتنی بر نرم

رایانش از جمله نظریه فازی و یادگیری ماشینی، یادگیری

الکترونیکی و دولت الکترونیک است.

نشانی رایانه ایشان عبارت است از:

Montazer@Modares.ac.ir