



دانشکده مهندسی کامپیوتر

پروژه مقطع کارشناسی مهندسی کامپیوتر

حذف انعکاس دوربین از شیشه در خودروهای خودران

سید شایان دانشور

استاد راهنما:

دکتر سید بهروز نصیحت کن

تابستان ۱۴۰۱

تأییدیه هیات داوران

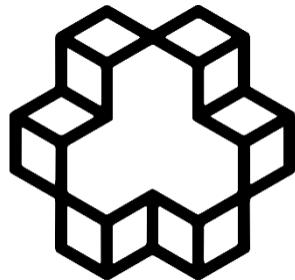
اعضای هیئت داوران، نسخه نهائی پروژه آقای: سید شایان دانشور

را با عنوان: حذف انعکاس دوربین از شیشه در خودرو های خودران

از نظر شکل و محتوی بررسی نموده و پذیرش آن را برای تکمیل درجه کارشناسی تأیید می کنند.

اعضای هیئت داوران	نام و نام خانوادگی	رتبه علمی	امضاء
۱- استاد راهنمای	سید بهروز نصیحت کن	استاد دیار	
۲- استاد داور	بابک ناصرشریف	استاد دیار	

ز



دانشگاه صنعتی خواجه نصیرالدین طوسی

اظهارنامه دانشجو

اینجانب سید شایان دانشور دانشجوی مقطع کارشناسی رشته مهندسی کامپیوتر گواهی می‌نمایم
که مطالب ارائه شده در این پروژه با عنوان:

حذف انعکاس دوربین از شیشه در خودروهای خودران

با راهنمایی استاد محترم سید بهروز نصیحت کن توسط شخص اینجانب انجام شده است. صحت و
اصالت مطالب نوشته شده در این پروژه تأیید می‌شود و در تدوین متن پروژه قالب مصوب دانشگاه را به
طور کامل رعایت کرده‌ام.

خالد
Shaygan

22/4/1401

امضاء دانشجو:

تاریخ:

حق طبع، نشر و مالکیت نتایج

- ۱- حق چاپ و تکثیر این پژوهه متعلق به نویسنده و استاد راهنمای آن است. هرگونه تصویربرداری از کل یا بخشی از پژوهه تنها با موافقت نویسنده یا استاد راهنمای یا کتابخانه دانشکده‌های مهندسی برق و کامپیوتر دانشگاه صنعتی خواجه نصیرالدین طوسی مجاز است.
- ۲- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی خواجه نصیرالدین طوسی است و بدون اجازه کتبی دانشگاه قابل واگذاری به شخص ثالث نیست.
- ۳- استفاده از اطلاعات و نتایج موجود پژوهه بدون ذکر مرجع مجاز نیست.

ط

تقدیم به:

پدر و مادر عزیز و مهربانم که در سختی ها و ناهمواری های زندگی هموار یاوری دلسوز و فداکار و پشتیبانی محکم و مطمئن برایم بوده اند.

ک

تشکر و قدردانی

بدینوسیله از زحمات فراوان استاد راهنمای محترم، جناب آقای دکتر سید بهروز نصیحت کن که در طول
فرایند انجام پایان نامه از راهنمایی ها و نظرات ارزشمند ایشان بهره بردم، قدردانی مینمایم.

چکیده

در سال های اخیر شاهد پیشرفت های بسیاری در حوزه های بینایی ماشین و پردازش تصویر و به سبب آن در حوزه های مرتبط از جمله خودروهای خودران بوده ایم. یکی از وظایف خودروهای خودران، تشخیص سریع و درست اشیا موجود در جاده است، که خوشبختانه امروزه با وجود مجموعه داده های بسیار، پیاده سازی چنین سیستم هایی با شبکه های عصبی بسیار ساده شده است. اما مشکل اساسی که وجود دارد این است که معمولاً این سیستم های تشخیص اشیاء تصویر مورد استفاده شان از دوربینی می آید که پشت شبشه جلوی خودرو نصب می شود و این باعث می شود انعکاس اشیاء داخل خودرو روی شبشه جلوی خودرو بیفتند و در کار الگوریتم های تشخیص اختلال ایجاد کند، چرا که این الگوریتم ها برای شرایط ایده آل نوشته شده اند و مدل های یادگیری عمیق مورد استفاده نیز روی مجموعه داده هایی که از بیرون خودرو عکس گرفته اند، آموزش می بینند. از طرف دیگر، از آنجایی که وضعیت قرارگیری شبشه جلوی خودرو، اشیا موجود درون خودرو و بازتاب آن روی شبشه جلو در هر خودرو متفاوت می شود، نیاز هست سیستمی به صورت مجزا داشته باشیم تا ابتدا انعکاس اشیا را از شبشه حذف کند و سپس تصویر را به الگوریتم های تشخیص بدهیم. در این پایان نامه، ابتدا مجموعه داده مورد نیاز که شامل تصویر حقیقی و تصویر همراه با بازتاب اشیاء باشد، ساخته شده و دو مدل شبکه عصبی مبتنی بر U-Net چهت یادگیری چگونگی حذف بازتاب ارائه شده است، هر دو مدل به دقت بالای ۷۱ درصد رسیده اند و عملکرد آن ها بررسی شده است. در نهایت پیشنهادهایی چهت بهبود مدل های مذکور ارائه شده و مدل جدیدی بر اساس نتیجه گیری های انجام شده پیشنهاد شده است.

کلید واژه: حذف انعکاس تصویر، جدا سازی لایه ها، بهبود تصویر، حذف انعکاس ویدیو، حذف انعکاس تک تصویر، U-Net

فهرست مطالب

صفحه		عنوان
۵	تأثیردیّه هیات داوران
۵	فهرست شکل‌ها
ز	فهرست جدول‌ها
۱	فصل ۱ - مقدمه
۱	۱-۱ پیشگفتار
۲	۲-۱ حذف بازتاب اشیاء از تصویر
۴	۴-۱ هدف و ساختار پایان نامه
۵	فصل ۲ - پیش زمینه
۵	۵-۱ مقدمه
۵	۵-۲ تصویر
۶	۶-۳ شبکه عصبی
۷	۷-۳-۲ معماری شبکه عصبی
۹	۹-۲ شبکه عصبی پیچشی
۱۰	۱۰-۲ لایه‌های پیچشی
۱۱	۱۱-۲ ادغام
۱۱	U-Net -۴-۲
۱۲	۱۲-۴-۲ قسمت رمزگذار
۱۳	۱۳-۴-۲ قسمت پل
۱۳	۱۳-۴-۲ قسمت رمزگشا
۱۴	۱۴-۴-۲ ارتباطات همراه با پرش
۱۵	فصل ۳ - پیاده سازی -۴-۲
۱۵	۱۵-۳ مجموعه داده ها
۱۵	۱۵-۱-۳ SIRR مجموعه داده
۱۶	۱۶-۱-۳ CamVid مجموعه داده

۱۷	تولید مجموعه داده از روی CamVid	-۳-۱-۳
۱۷	مجموعه داده حقیقی همراه با بازتاب طبیعی	-۴-۱-۳
۲۰	تقسیم بندی داده های موجود	-۵-۱-۳
۲۱	روش های پیاده سازی شده	-۲-۳
۲۱	روش اول	-۱-۲-۳
۲۲	روش دوم	-۲-۲-۳
۲۲	پیش پردازش داده های ورودی	-۳-۲-۳
۲۲	آموزش	-۴-۲-۳
۲۳	فصل ۴ - ارزیابی، نتیجه گیری و پیشنهادها	
۲۳	ارزیابی	-۴-۱
۲۳	دقت	-۱-۱-۴
۲۴	داده آموزش	-۲-۱-۴
۲۵	داده تست	-۳-۱-۴
۲۷	داده حقیقی با بازتاب طبیعی	-۴-۱-۴
۲۸	نتیجه گیری	-۲-۴
۲۸	پیشنهادها	-۳-۴
۳۱	پیوست ۵ - واژه‌نامه فارسی-انگلیسی	
۳۳	پیوست ۶ - واژه‌نامه انگلیسی-فارسی	
۳۶	فهرست مرجع‌ها	

فهرست شکل‌ها

صفحه

عنوان

۷	شکل ۱-۲ شبکه عصبی با یک لایه میانی.....
۱۰	شکل ۲-۲ ساختار یک شبکه عصبی پیچشی برای کلاس بندی حیوانات.....
۱۲	شکل ۲-۳-۲ معماری شبکه U-NET.....
۱۶	شکل ۳-۱ نمونه تصویر حقیقی از SIRR.....
۱۶	شکل ۳-۲ نمونه تصویر با بازتاب از SIRR.....
۱۷	شکل ۳-۳ نمونه تصویر از مجموعه داده CAMVID.....
۱۸	شکل ۴-۱ نمونه اول تصویر ساخته شده از مجموعه داده CAMVID.....
۱۹	شکل ۴-۲ نمونه دوم تصویر ساخته شده از مجموعه داده CAMVID.....
۱۹	شکل ۴-۳ نمونه اول تصویر بازتاب.....
۱۹	شکل ۴-۴ نمونه دوم تصویر بازتاب.....
۲۴	شکل ۴-۱ نمونه تصویر آموزش و عملکرد شبکه ها روی آن.....
۲۵	شکل ۴-۲ نمونه تصویر تست و عملکرد شبکه ها روی آن.....
۲۶	شکل ۴-۳ نمونه تصویر تست و عملکرد شبکه ها روی آن.....
۲۷	شکل ۴-۴ نمونه تصویر واقعی با بازتاب طبیعی - جاده خالی.....
۲۷	شکل ۴-۵ نمونه تصویر واقعی با بازتاب طبیعی - جاده شلوغ.....

فهرست جداول

عنوان	صفحه
جدول ۱-۴ نتایج ارزیابی دو شبکه بر روی داده تست.....	۲۳

فصل ۱- مقدمه

۱-۱- پیشگفتار

امروزه تولید و استفاده از خودرو های خودران و بسیار فراگیر شده است. خودرو های خودران وظایف بسیار و پیچیده ای دارند که مهم ترین آن ها امکان تشخیص اشیائی^۱ همچون عابران پیاده، خطوط جاده، تابلو های راهنمایی و موانع می باشد. چنین وظیفه ای به دلیل پیچیدگی زیاد، با استفاده از تکنیک های پردازش تصویر^۲، بینایی کامپیوتر^۳، یادگیری ماشین^۴ و یادگیری عمیق^۵ صورت می گیرد. پیاده سازی سیستم های تشخیص اشیاء، به این صورت انجام می شود که از مجموعه داده های^۶ آماده ای که حاوی تصویر شی و نام آن تصویر است صورت می گیرد. این تصاویر معمولا در شرایط نسبتاً ایده آلی گرفته شده اند و در آن ها اثری از گرد و غبار، قطرات باران و نیز بازتاب اشیاء نیست. بنابرین در صورت استفاده از سیستمی که روی چنین مجموعه داده ای آموزش دیده، باید حواسمن باشد که در صورت وجود گرد و غبار بیش از حد و یا بازتاب اشیاء دیگر روی شی مورد نظر، چندان خوب عمل نمی کنند. یکی از مشکلات پیش روی خودرو های خودران نیز این است که دوربین، پشت شیشه جلو نصب شده و انعکاس اشیاء داخل خودرو با شدت زیادی بر روی شیشه جلوی خودرو و در نتیجه درون تصویر گرفته شده توسط دوربین می افتد و مانع کار کرد درست سیستم

^۱ Object Detection

^۲ Image Processing

^۳ Computer Vision

^۴ Machine Learning

^۵ Deep Learning

^۶ Dataset

ها و الگوریتم های تشخیص می شوند. بنابرین وجود سیستمی که بتواند ابتدا بازتاب را از شیشه حذف کند یا حداقل اثر آن را کم کند به شدت ارزشمند و مورد نیاز است.

۱-۲- حذف یا زتاب اشیاء از تصویر

تصاویر حاوی بازتاب اشیاء را می توان ترکیبی خطی از دو لایه تصویر در نظر گرفت و با رابطه (۱-۱) نشان داد.

$$I = w \times T + (1 - w) \times (K * R) + n \quad (1-1)$$

که در آن I تصویر نهایی، T تصویر مربوط به بخش اصلی تصویر، R تصویر اشیاء بازتاب شده پس زمینه، K یک فیلتر تار کننده گاوسی^۱، n نویز به وجود آمده هنگام عکس برداری و w ضریبی بین صفر و یک می باشد. واضح است که مسئله جدا سازی R از I کاری بسیار دشوار می باشد، چرا که تنها یک معادله و چندین مجھول داریم، به طوریکه حتی اگر تصویر نهایی را ترکیب ساده T و I در نظر بگیریم و فرض کنیم که ضریب w در معادله وجود ندارد و نویز را نیز صفر فرض کنیم، آنگاه همچنان مسئله دشواری برای حل داریم چراکه یک معادله و دو مجھول داریم:

روش های حذف بازتاب اشیاء از تصاویر، از جهت تعداد تصاویر، به دو دسته تک تصویره و چند تصویره تقسیم می شود، که اکثر تحقیقات انجام شده تا به امروز بر روی تک تصویر بوده است.

برای حذف بازتاب اشیاء از تصویر، روش های کلاسیک [۲، ۳، ۴] به طور عمدۀ با تمرکز بر ویژگی تصاویر حقیقی و نیز ویژگی بازتاب برای حذف استفاده می کنند. یکی از پر استفاده ترین این روش ها، استفاده از ویژگی آماری تصاویر حقیقی [۱] و برازش یک یا دو توزیع بر روی تصویر است، به این صورت که می دانیم که گرادیان تصاویر طبیعی تنک^۲ هستند و از طرفی فرض می شود که توزیع آماری لایه اصلی و بازتاب متفاوت بوده و هیچ همبستگی با یکدیگر ندارند. به این ترتیب با برازش توزیعی شامل دو توزیع بر روی داده

Blurring Gaussian Filter

Sparse

و تلاش برای جدا سازی دو لایه از طریق بهینه سازی دو لایه را جدا می کنند. روش های کلاسیک دیگر شامل کلاس بندی لبه ها از طریق میزان نرمی آن ها و سپس جدا سازی لایه ها [۵]، تشخیص و جداسازی دو لایه با فرض تفاوت میزان کدری لایه ها [۶]، مقایسه دو تصویر گرفته شده با و بدون فلاش [۷]، دنبال کردن حرکت اشیاء در دنباله ای از تصاویر با استفاده از روش هایی همچون شار نوری^۱ و یا استفاده از فیلتر های فضایی-زمانی و کلاس بندی دو لایه [۸، ۹، ۱۰، ۱۱، ۱۲] و نیز تشخیص بازتاب با فرض بر تکرار شدن آن در شیشه های ضخیم [۱۳] می باشد. در همه این روش ها، فرض هایی بر روی چیستی و چگونگی به وجود آمدن بازتاب گذاشته می شود و هنگام تست شدن روی تصاویر واقعی همراه با بازتاب طبیعی، عملکرد چندان قابل قبولی نشان نمی دهدند مگر در شرایط خاصی که فرض های گرفته شده کاملاً صادق باشند.

در روش های جدیدتر که مبتنی بر یادگیری ماشین و به خصوص یادگیری عمیق می باشند، نیز مشکل مجموعه داده وجود دارد، چرا که در حال حاضر هیچ مجموعه داده ای حاوی تصاویر واقعی با و بدون بازتاب طبیعی وجود ندارد و مدل های مبتنی بر یادگیری عمیق تنها در صورتی به خوبی عمل می کنند که مجموعه داده خوب و بزرگی برای آموزش آن داشته باشیم. بنابرین در روش های ارائه شده، برای ساخت مجموعه داده از تصاویر طبیعی استفاده شده و سپس بازتاب به صورت مصنوعی روی این تصاویر قرار گرفته است، به این صورت که تصویر طبیعی دومی، پس تار شدن و نیز کاهش شدت رنگ آن، در قسمت یا قسمت هایی از تصاویر قرار گرفته اند. بنابرین، همانطور که انتظار می رود، چنین مدل هایی نیز عملکرد عالی ندارند، چون مجموعه داده آن ها ساختگی است و معمولاً با دنیای واقعی تفاوت زیادی دارد، به این معنی که اگر همان تصاویر ساختگی را به مدل بدهیم، بازتاب مصنوعی به خوبی از تصویر حذف می شود ولی اگر یک تصویر واقعی همراه با بازتاب طبیعی به مدل بدهیم، عملکرد چندان قابل قبولی در شرایط گوناگون از خود نشان نمی دهد. ولی اگر بازتاب مصنوعی به وجود آمده روی مجموعه داده ای مشابه فضایی که میخواهیم در آن

^۱ Optical Flow

فضا از مدل یادگیری عمیق استفاده کنیم باشد، می‌توان تا حدودی انتظار داشت که مدل بتواند بازتاب را تا حد خوبی و فقط برای فضای مشابه مجموعه داده حذف کند.

نمونه‌هایی از مدل‌های عمیق ارائه شده شامل، استفاده از شبکه‌های عمیق پیچشی^۱ [۱۴]، استفاده از شبکه‌های عمیق رمزگذار-رمزگشای^۲ [۱۵]، استفاده از شبکه‌های مولد رقابتی^۳ [۱۶] و نیز استفاده از شبکه‌های حافظه کوتاه-مدت ماندگار^۴ [۱۷] می‌باشد.

۱-۳-۱ هدف و ساختار پایان نامه

هدف این پایان نامه ارائه و پیاده سازی روشی جهت حذف بازتاب دوربین از تصویر در خودروهای خودران است، به این معنی که صرفا، هدف ما این است که بازتاب را از تصاویری که از درون خودرو گرفته شده اند حذف کنیم.

در ادامه این پایان نامه، ابتدا در فصل دوم به بیان مفاهیم و اصطلاحات ضروری و مهم می‌پردازیم، سپس در فصل سوم به روش پیشنهادی جهت حذف بازتاب و توضیح پیاده سازی می‌پردازیم و مفصل آن را بررسی می‌کنیم. در نهایت در فصل چهارم، نتایج بدست آمده را بررسی و مقایسه و نتیجه گیری می‌کنیم و در آخر پیشنهادهایی را جهت بهبود روش مذکور ارائه می‌دهیم.

^۱ Deep Convolutional Network

^۲ Encoder-Decoder

^۳ Generative Adversarial Network

^۴ Long Short-Term Memory Network

فصل ۲ - پیش زمینه

۱-۲ مقدمه

در این فصل به ارائه مفاهیم اولیه و مورد نیاز که در پروژه استفاده شده اند می پردازیم. در ابتدا به تعریف تصویر پرداخته و ساختار آن در کامپیوتر را بررسی می کنیم، سپس به شبکه عصبی^۱ و شبکه عصبی پیچشی و مفاهیم آن می پردازیم، در نهایت به معرفی و توضیح شبکه عصبی پیچشی موسوم به U-Net می پردازیم.

۲-۲ تصویر

در پردازش تصویر و بینایی کامپیوتر، تصویر به شکل یک آرایه در حافظه ذخیره می شود، در تصاویر سیاه سفید، هر عضو این آرایه بیانگر یک پیکسل از تصویر می باشد و مقدار آن میزان روشنایی یا سفید بودن آن پیکسل را بیان می کند. برای تصاویر رنگی روش ها و مدل های مختلفی برای ذخیره سازی وجود دارد که متداول ترین آن ها RGB می باشد. در این حالت، به جای یک عدد، سه عدد بیانگر یک پیکسل می باشند و این سه عدد، بیانگر میزان روشنایی رنگ های قرمز، سبز و آبی هر پیکسل می باشند. در این مدل رنگ، شدت رنگ ها بین ۰ تا ۲۵۵ قرار دارد.

^۱ Neural Network

۳-۲ شبکه عصبی

به طور کلی، در الگوریتم های یادگیری ماشین نیاز داریم تا ویژگی^۱ هایی را از داده های خام استخراج کنیم. این ویژگی ها می توانند ساده باشند و مستقیماً از داده خام به عنوان ویژگی استفاده شده باشند و یا می توانند پیچیده بوده و از حاصل عملیات های پیچیده ریاضی، آماری و سیگنالی بدست آمده باشند. برای مثال یک ماشین بردار پشتیبان^۲ ساده که از قدیمی ترین الگوریتم های یادگیری ماشین است را در نظر بگیرید که قرار باشد که با گرفتن یک عکس به عنوان ورودی، مشخص کند که این عکس تصویر یک سیب است یا پرتقال. ماشین بردار پشتیبان در ورودی، همواره یک بردار دریافت می کند. این بردار در ساده ترین حالت، می تواند تصویر خام و شدت روشنایی پیکسل های آن باشد. اما، باید توجه داشت که ورودی خام گزینه مناسبی به عنوان ویژگی نمی باشد و ماشین بردار پشتیبان به درستی آموزش^۳ نمی بیند و دقت خوبی ارائه نخواهد داد. از این رو برای این مسئله باید ویژگی های پیچیده تر و سطح بالاتری استخراج شود و به مدل داده شود تا آموزش ببیند. امروزه در کاربرد های پیچیده، انتخاب ویژگی مناسب به سادگی امکان پذیر نیست و محققان را به مجبور به روی آوردن به روش های دیگر جهت استخراج ویژگی کرده است. در چنین مسائلی که استخراج ویژگی به صورت دستی امکان پذیر نیست، شبکه های عصبی می توانند راه حل خوبی باشند. شبکه های عصبی، با الهام گیری از مدل مغز انسان طراحی شده اند. به عنوان نمونه، شبکه های عصبی پیچشی تلاش می کند با استفاده از استخراج ویژگی های ساده و استفاده از آن ها به ویژگی های پیچیده تری دست یابد. در این نوع شبکه ها، با افزایش عمق شبکه، می توان ویژگی های پیچیده تری را استخراج کرد و مدل خود را با دقت و کارایی بالاتری آموزش داد. با توجه به اینکه شبکه های عصبی، ویژگی ها را به صورت خودکار استخراج می کنند، می توان آن ها را یکی از قدرتمند ترین ابزار های یادگیری ماشین به حساب آورد.

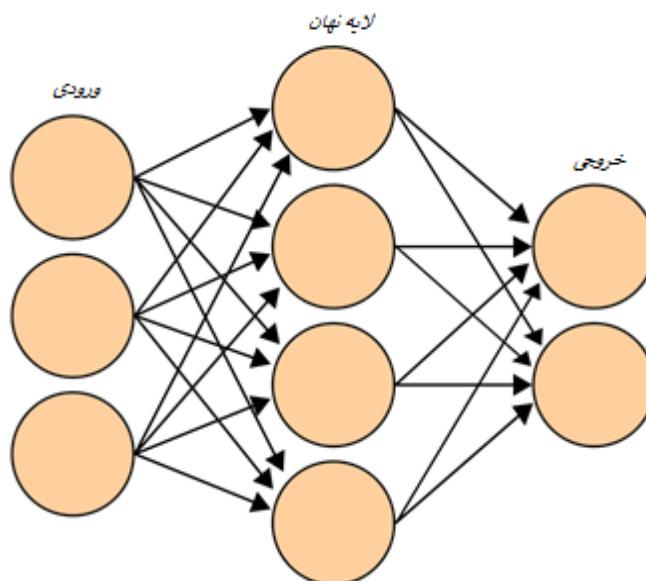
¹ Features

² Support Vector Machine

³ Train

۳-۲-۱-معماری شبکه عصبی

به طور کلی هر شبکه عصبی از یک لایه ورودی، چند لایه میانی یا مخفی و یک لایه خروجی تشکیل شده است. هر لایه از تعدادی نورون^۱ تشکیل می شود. هر نورون با انجام عملیاتی مشخص، روی ورودی های خود یک خروجی را محاسبه می کند. شکل ۱-۲ یک نمونه ساده از شبکه عصبی با یک لایه میانی را نشان می دهد.



شکل ۱-۲: شبکه عصبی با یک لایه میانی [۲۳]

همانطور که در شکل ۱-۲ مشاهده می شود، هر نورون با یک دایره مشخص شده و به تمامی نورون های لایه های قبل و بعد از خود متصل است. این نوع شبکه عصبی را شبکه عصبی را شبکه عصبی تماماً متصل^۲ می گویند. همچنین هر اتصال، معادل یک مقدار است که به آن وزن^۳ می گویند. به طور کلی هر شبکه عصبی دارای پارامتر^۴ هایی است و هدف آن این است که با تنظیم مقدار این پارامتر ها بهترین نتیجه را در خروجی داشته باشد. به این فرایند یادگیری می گویند، چرا که شبکه سعی می کند بهترین مقادیر برای پارامتر ها را با ارزیابی خروجی شبکه، یاد بگیرد. به طور کلی، دو دسته پارامتر وزن و بایاس^۵ در شبکه وجود دارد و هر لایه

^۱ Neuron

^۲ Fully Connected Neural Network

^۳ Weight

^۴ Parameter

^۵ Bias

پارامترهای خود را دارد. در شبکه عصبی، هر نورون وظیفه دارد که پس از اعمال این دو پارامتر مطابق با رابطه (۱-۲) حاصل را به یک تابع غیر خطی به نام تابع فعال ساز وارد نماید، خروجی را محاسبه کند و آن را به لایه‌های بعد انتقال دهد.

$$y = f(x^T w + b) \quad (1-2)$$

در رابطه (۱-۲)، f تابع فعال ساز، X خروجی لایه قبل یا همان ورودی شبکه، W وزن‌های مرتبط با هر نورون، b مقدار بایاس و لاخروجی نورون می‌باشد. باید توجه داشت که ورودی و وزن هر دو بردارهای یک بعدی می‌باشند.

دلیل وجود تابع فعال ساز، این است که اگر هر لایه تنها حاصل جمع و ضرب مقادیر باشند، از آنجا که هر دو اپراتورهایی خطی هستند، توالی لایه‌های مختلف نیز خطی می‌شوند و کل شبکه را می‌توان تنها با یک لایه نمایش داد و پیاده سازی کرد، چنان شبکه‌ای تنها می‌توان روابط خطی بین ورودی و خروجی را یاد بگیرد. به همین دلیل تابع فعال ساز تعریف می‌شود که معمولاً تابع غیر خطی است و در اثر وجود آن است که ویژگی‌های غیر خطی می‌توانند استخراج شوند.

روال یادگیری شبکه‌های عصبی نیز، معمولاً به این صورت است که پس از فاز انتشار رو به جلو^۱ که همان مقدار دهی شبکه و محاسبه مقادیر نورون‌ها و در نهایت دادن خروجی نهایی می‌باشد، یک تابع هزینه تعریف می‌شود که تفاوت مقدار مورد نظر، با مقدار پیش‌بینی شده توسط شبکه را حساب می‌کند. سپس این خطای عملیاتی که از آن تحت عنوان پس انتشار^۲، یاد می‌شود به لایه‌های قبلی باز میگردد و خطای هر نورون

^۱ Forward Propagation

^۲ Back Propagation

محاسبه می گردد. سپس این مقادیر به یک الگوریتم بهینه سازی، همچون گرادیان نزولی^۱ داده می شود تا تعیین کند پارامتر های هر نورون چه میزان تغییر کنند تا در تکرار^۲ بعدی و باز مقدار دهی شبکه، نتیجه بهتری بگیریم. در واقع، طی روال یادگیری شبکه،تابع هزینه بهینه می شود و تا حد امکان به صفر نزدیک می شود.

-۲-۳-۲ شبکه عصبی پیچشی

با گسترش شبکه های عصبی عمیق تماما متصل و پی بردن به قدرت آن ها در استخراج ویژگی، محققان کوشیده اند تا از آن در حوزه های گوناگون استفاده کنند و بسته به کاربرد، معماری های جدیدی را پیشنهاد دهند. یکی از این حوزه ها، بینایی کامپیوتر و پردازش تصویر است که شبکه های عصبی تماما متصل در آن ها نسبتا ضعیف عمل می کنند و برای رسیدن به دقت مناسب، نیاز است شبکه ای به شدت عمیق داشته باشیم که آموزش آن ممکن نیست، چرا که با عمیق شدن شبکه ها، مشکلی موسوم به گم شدن گرادیان ها^۳ رخ می دهد. اما دلیل ضعف شبکه عصبی معمولی برای کار های مختلف مرتبط با تصویر، این است که در تصویر، هر پیکسل به تنها یک اطلاعات خوبی نمی دهد و ویژگی چندان مناسبی نیست و هر پیکسل به همراه پیکسل های اطراف خود معنی پیدا می کند و این ویژگی های محلی^۴ تصویر است که مشخص میکند در یک قسمت از تصویر چه چیزی قرار دارد.

در یک تصویر علاوه بر مقدار هر پیکسل، پیکسل های مجاور و نحوه چینش آن ها در کنار یکدیگر نیز اهمیت دارد. برای مثال پیکسل ها در کنار یکدیگر ممکن است یک گوشه و یا لبه ای را تشکیل دهند و این ویژگی حائز اهمیت باشد. با پیدایش شبکه های عصبی پیچشی، این مشکل تا حد قابل توجهی برطرف شد.

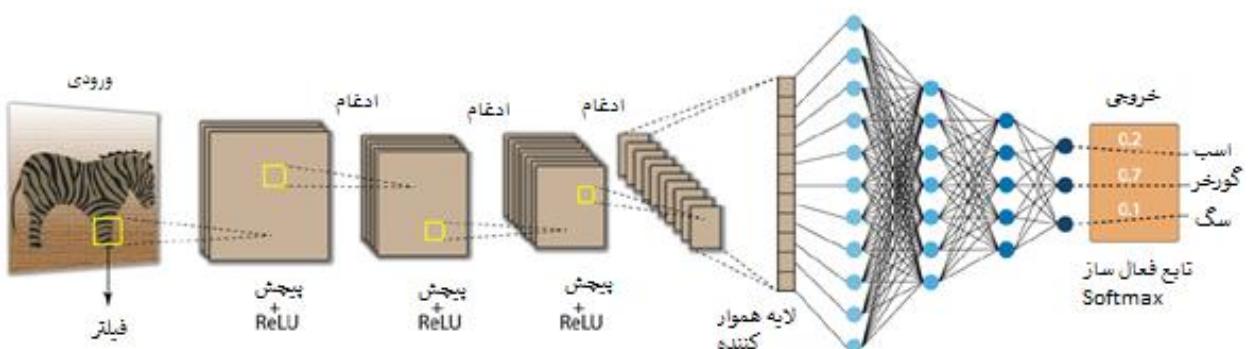
^۱ Gradient Descent

^۲ Iteration

^۳ Vanishing Gradients Problem

^۴ Spatial Features

شبکه های عصبی پیچشی، به جای استفاده از عملیات جمع و ضرب، از عملیات خطی پیچش^۱ استفاده میکند، به این صورت که وزن ها در قالب یک فیلتر^۲ روی تصویر پیمایش می شوند و با اعمال عملیات پیچش و عبور دادن آن از تابع فعال ساز در هر قدم مقادیر جدیدی محاسبه و ذخیره می شوند. از این رو یک فیلتر بعد از پیمایش کل تصویر، تصویر جدیدی را بدست می آورد که هر پیکسل آن می تواند بیانگر ویژگی جدیدی باشد. از مزایا و البته محدودیت های این نوع شبکه این است که پارامتر های هر لایه بر خلاف شبکه عصبی تماماً متصل ارتباطی به ورودی ندارند و مقداری ثابت هستند، از طرف دیگر فیلتر های لایه های مختلف به یکدیگر وابسته نیستند و می توان آن ها را با صورت موازی روی تصویر اعمال کرد که بازدهی و سرعت یادگیری را بالا می برد. نکته مهم در مورد شبکه های عصبی پیچشی این است که عموماً در معماری شبکه های مبتنی بر آن ها، علاوه بر لایه های پیچشی، لایه های تماماً متصلی قرار میگیرند و اطلاعات استخراج شده توسط لایه های پیچشی را تحلیل کرده و خروجی می دهند، اما این به این معنی نیست که شبکه های تماماً پیچشی^۳ کاربرد ندارند. نمونه ای از یک شبکه عصبی پیچشی همراه با یک شبکه عصبی تماماً متصل در شکل (۲-۲) آمده است.



شکل (۲-۲): ساختار یک شبکه عصبی پیچشی برای کلاس بندی حیوانات [۲۴]

^۱ Convolution

^۲ Filter

^۳ Fully Convolutional Network

-۳-۳-۲ لایه های پیچشی

لایه اصلی در هر شبکه عصبی پیچشی است. بیشتر محاسبات و استخراج ویژگی در این لایه انجام می شود. هر لایه پیچشی شامل تعدادی فیلتر است و خروجی هر لایه حاصل عمل پیچش میان فیلتر ها و ورودی آن لایه است. هدف از وجود تعداد زیادی فیلتر، یادگیری و استخراج ویژگی های گوناگون تصویر بوده و تعداد این فیلتر ها معمولاً توانی از دو بوده و اندازه فیلتر ها معمولاً عددی فرد بوده و معمولاً طول و عرض آن با هم برابر است. معمولاً با افزایش تعداد فیلتر ها، شبکه قدرتمند تر شده و قادر خواهد بود مسائل پیچیده تری را حل کند.

-۴-۳-۲ ادغام

هدف لایه ادغام کاهش اندازه خروجی به دست آمده از لایه پیچشی است. لایه ادغام پارامتر قابل آموزش ندارد و یک نمونه برداری ساده انجام می دهد. فرایнд ادغام از این جهت مشابه پیچش است که یک فیلتر یا پنجره روی تصویر حرکت می کند. رایج ترین نمونه های آن ادغام میانگین و ادغام حداکثری بوده و اندازه خروجی به تنظیمات که برای آن در نظر گرفته می شود دارد، معمولاً اندازه خروجی نصف لایه قبل آن می شود. معمولاً از ادغام میانگین پس از آخرین لایه پیچشی و از ادغام حداکثری در لایه های میانی قبل از آن استفاده می شود.

U-Net -۴-۲

U-Net یک شبکه عصبی تماماً پیچشی است که اولین بار برای بخش بندی تصویر^۱ از آن استفاده شده [۲۰] و در آن از چهار مفهوم قسمت های رمز گذار^۲ و رمزگشا^۳، قسمت پل^۴ و ارتباطات همراه با پرش^۵ استفاده شده است. به این صورت که معماری شبکه را می توان شامل سه قسمت که قسمت اول رمزگذار، قسمت

^۱ Image Segmentation

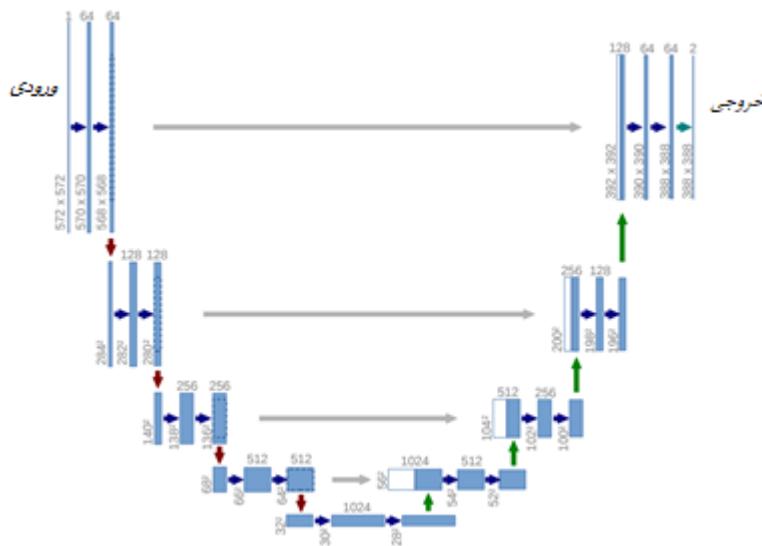
^۲ Encoder

^۳ Decoder

^۴ Bridge

^۵ Skip Connections

میانی و دوم پل و قسمت سوم و نهایی را رمزگشا دانست که ارتباطاتی بین قسمت رمزگذار و رمزگشا وجود دارد که در ادامه بیشتر به توضیح هر قسمت می‌پردازیم. معماری این شبکه در شکل (۳-۲) آمده است.



شکل (۳-۲): معماری شبکه U-Net [۲۰]

همانطور که در شکل (۳-۲) دیده می‌شود معماری رسم شده، به شکل U در آمده و دلیل نام‌گذاری آن نیز همین می‌باشد.

۱-۴-۲- قسمت رمزگذار

این قسمت شامل هشت لایه پیچشی با اندازه فیلتر ۳ در ۳ می‌باشد که پس از هر لایه پیچش، یک لایه ReLU [۲۱] و یک لایه BatchNormalization از دو لایه یکبار از یک لایه ادغام حداقل با اندازه فیلتر ۲ در ۲ استفاده شده است. ReLU یک تابع فعال ساز معروف و بسیار پر کاربرد در شبکه‌های عصبی پیچشی است که در آن ورودی با مقدار صفر مقایسه می‌شود و مقدار بزرگتر به خروجی داده می‌شود، به عبارت دیگر $ReLU(x) = \max(0, x)$. منظور از لایه Normalization که خود نیز بسیار شناخته شده بوده و در اکثر مواقع در کنار تابع فعال ساز استفاده می‌شود، تکنیکی است که به پایداری شبکه و یادگیری بهتر آن کمک می‌کند.

هدف از این قسمت، استخراج ویژگی قسمت های مختلف تصویر می باشد. واضح است که در این قسمت به دلیل استفاده از فیلتر های پیچشی همراه با فیلتر های ادغام ابعاد تصویر کوچک می شود و بخش هایی از اطلاعات تصویر از بین می رود.

-۴-۲ قسمت پل

این قسمت تنها شامل دو لایه پیچشی مشابه قسمت رمزگذار است ولی فاقد لایه ادغام می باشد و تنها پلی بین دو قسمت ابتدایی و انتهایی شبکه می باشد. در این قسمت اندازه تصویر حفظ می شود و تنها بخش کوچکی که در واقع پیکسل های دور تا دور تصویر هستند به دلیل لایه های پیچشی از بین می روند که در این لایه ها حذف شان نمی کنیم بلکه با مقدار صفر دادن، اندازه تصویر را حفظ می کنیم.

-۴-۳ قسمت رمزگشایی

این قسمت شامل ۴ لایه پیچش ترانهاده^۱ است که پس از هر لایه، یک لایه الحقی^۲ و دو لایه پیچشی ReLU و پس از هر لایه پیچش مشابه قسمت رمزگذار، یک لایه Batch Normalization و یک لایه فیلتر ضرب قرار دارد. منظور از پیچش ترانهاده عکس عمل پیچش نیست ولی تا حدودی مشابه آن عمل میکند، به طوریکه که فیلتر پیچش ترانهاده، هر کدام از پیکسل های موجود درون خود را در همه خانه های فیلتر ضرب کرده و به تعداد اعضای داخل فیلتر، فیلتر به وجود می آورد و در نهایت این فیلتر هارا کنار یکدیگر می گذارد و در قسمت هایی که فیلتر ها روی یکدیگر می افتدند مقادیر آن ها را با یکدیگر جمع می کنیم، به طوریکه در خروجی تصویر بزرگتری خواهیم داشت و به لحاظ تغییر اندازه تصویر طوری عمل می شود که اندازه تصویر به اندازه تصویر قبل از پیچش مشابه باز می گردد. در این قسمت به دلیل استفاده از لایه های پیچش ترانهاده، اندازه تصویر بزرگتر می شود. لازم به ذکر است که پیچش ترانهاده تنها روش بازگردانی تصویر به ابعاد بزرگتر نیست و روش های دیگری نیز وجود دارد.

^۱ Transposed Convolution

^۲ Concatenation

هدف از این لایه تلاش برای بازسازی قسمت‌های از دست رفته در قسمت رمزگذار است، در واقع در این قسمت شبکه سعی می‌کند، اطلاعات از دست رفته در قسمت رمزگذار را به گونه‌ای دیگر بازسازی کند و به جهت تسهیل روند بازسازی و نیز جلوگیری از مشکل گم شدن گرادیان‌ها، اطلاعات از لایه‌های قبلی در کنار اطلاعات لایه فعلی قرار می‌گیرند و الحق می‌شوند. بنابرین U-Net علاوه بر توانایی بازسازی قسمت‌هایی از تصویر، مشکل گم شدن گرادیان‌ها را نیز حل می‌کند و باعث می‌شود بتوانیم لایه‌های شبکه را بسته به کاربرد تا حد قابل قبولی افزایش دهیم. به سبب همین ویژگی هاست که U-Net در بسیاری از کاربردها از جمله ترجمه تصویر به تصویر^۱ [۲۲] و بازسازی قسمت‌های از بین رفته تصویر استفاده شده‌اند.

در نهایت، در آخرین لایه قسمت رمزگشا، یک لایه پیچش نهایی با تابع فعال ساز معروف سیگموید قرار دارد که خروجی آن همواره بین صفر و یک است، دلیل استفاده از این تابع فعال ساز این است که معمولاً قبل کار کردن با تصویر در شبکه‌های عصبی، مقادیر پیکسل‌ها با تقسیم بر ۲۵۵ نرمال می‌شود و دیگر نیازی به توابع فعال سازی که مقداری بیش از ۱ می‌دهند نیست. در چنین شرایطی به تجربه به خوبی دریافته شده است که شبکه با تابع فعال ساز سیگموید در آخرین لایه، زودتر از شبکه مشابه با تابع فعال سازی که بین صفر و یک نیست و فراتر از آن است، آموزش می‌بینند و همگرا می‌شود.

۴-۴-۲ ارتباطات همراه با پرس

منظور از ارتباطات همراه با پرس، در واقع همان لایه‌های الحاقی موجود در بخش رمزگشا است که اطلاعات همان لایه با اطلاعات لایه مقابل در قسمت رمزگذار در کنار یکدیگر قرار می‌گیرند و روند دیکد کردن و بازسازی داده را تسهیل می‌کنند. ارتباطات همراه با پرس خاص U-Net نبوده و پیش از آن، در شبکه‌هایی همچون ResNet وجود داشته، اما در اکثر شبکه‌های پیش از آن، جلوگیری از مشکل گم شدن گرادیان‌ها بوده است.

^۱ Image-to-Image Translation

فصل ۳- پیاده سازی

۱-۳ مجموعه داده ها

متاسفانه در حال حاضر هیچ مجموعه داده مناسبی برای بازتاب اشیاء وجود ندارد و در تمام مقالات یا تصاویری به صورت مصنوعی تولید شده و ترکیبی از این تصاویر مصنوعی را به عنوان تصویر حاوی بازتاب در نظر گرفته اند و یا از تصاویر واقعی فاقد بازتاب استفاده کرده و تصویر دیگری را به صورت مصنوعی و دستی به تصویر واقعی اضافه کرده اند.

۱-۱-۳- مجموعه داده ^۱SIRR

معروف ترین مجموعه داده موجود با تصاویر واقعی و بازتاب مصنوعی، مجموعه داده SIRR [۱۸] می باشد. تعداد تصاویر موجود در این مجموعه داده ۱۰۸۱ تصویر می باشد که بر روی هر کدام از آن ها، بازتاب تصویر حقیقی دیگری به صورت مصنوعی اضافه شده است، بنابرین ۱۰۸۱ تصویر خام و ۱۰۸۱ تصویر حاوی بازتاب در تصویر وجود دارد. در شکل های (۱-۳) و (۲-۳) به ترتیب یک نمونه تصویر حقیقی و نیز تصویر حاوی بازتاب مصنوعی را می بینیم.

^۱ SIRR² نیز نوشته می شود



شکل (۳-۲): نمونه تصویر حقيقى از SIRR [۱۸]

مشکل اساسی مجموعه داده SIRR، اين است که اکثريت قریب به اتفاق تصاویر موجود در آن از محیط های داخلی و سر بسته گرفته شده و تعداد انگشت شماری از تصاویر آن در محیط خارج گرفته شده و اين تعداد نیز از فاصله نزدیک گرفته شده و عمق زیادی ندارند، که اين برخلاف تصاویر گرفته شده از درون خودرو است که عمق تصاویر معمولاً زیاد بوده و تا فاصله نسبتاً زیادی از جاده، درون تصویر دیده می شود.

مشکل ديگر آن مصنوعی بودن بازتاب ها می باشد که باعث می شود در صورت آموزش شبکه، در عین يادگيري چگونگي حذف بازتاب های مصنوعی، نتواند بازتاب های حقيقی را حذف کند، چرا که به لحظه ساختاري، بازتاب حقيقی بسته به محيط و شرایط ویژگی های متفاوتی می تواند داشته باشد و شبکه بازتاب های حقيقی را بخشی از تصویر اصلی در نظر بگیرد و قسمت هایی از تصویر که تاری بیشتری دارند را به جای بازتاب حذف کند.

۳-۱-۲-مجموعه داده CamVid

از آنجا که مجموعه داده SIRR به دلایلی که پیشتر ذکر شد، چندان قابل استفاده نیست و نمی توان تنها به آن متکی بود و از طرفی مجموعه داده مناسبی از محیط خارج که مشابه تصاویر گرفته شده از درون خودرو به سمت جاده باشد برای حذف بازتاب وجود ندارد، نیاز هست تا مجموعه داده مناسبی ساخته شود. برای

این کار از مجموعه داده CamVid [۱۹] استفاده می کنیم. این مجموعه داده، شامل تعدادی فیلم ضبط شده از حرکت خودرو در خیابان های انگلیس است که در کنار فیلم ها، تعدادی عکس از این ویدیو ها قرار دارد که هر یک ثانیه و با فرکانس یک هرتز از این ویدیو ها نمونه برداری شده اند و در نهایت تعداد ۷۰۱ تصویر بدست آمده اند، که می توان از آن ها برای تولید داده مورد نیاز استفاده کرد. نمونه ای از این مجموعه داده در شکل (۳-۳) آمده است.



شکل (۳-۳): نمونه تصویر از مجموعه داده CamVid

۳-۱-۳ - تولید مجموعه داده از روی CamVid

برای تولید داده جدید و افزودن تصاویر مصنوعی به صورت بازتاب به تصاویر، به طوری عمل شده است که از هر تصویر، ۱۶ تصویر حاوی بازتاب ساخته می شود. به این صورت که در ابتدا تصویر آینه شده تصویر اصلی ساخته می شود و یک تصویر تبدیل به دو تصویر می شود و سپس از روی هر تصویر، تعداد ۸ تصویر مصنوعی ساخته می شود. دلیل آینه کردن تصاویر این است که اولاً با آینه کردن افقی تصاویر، ضمن اینکه تصویر نسبتاً متفاوت ولی کاملاً طبیعی به وجود می آید و اندازه مجموعه داده عملاً دو برابر می شود، جهت

خیابان‌ها بر عکس می‌شود که بسیار مورد نیاز است، چرا که در این مجموعه داده، ویدیو‌ها درون کشور انگلستان ضبط شده‌اند و جهت رانندگی درون جاده‌های آن بر عکس ایران می‌باشد و آینه کردن باعث می‌شود تصاویر مشابه ایران درون مجموعه داده قرار گیرد.

پس از آینه کردن، حال روی هر تصویر بدست آمده به صورت تصادفی بین یک الی شش تصویر از پنج تصویر انتخابی از اینترنت انتخاب می‌شود که ممکن است در قسمتی کاملاً تصادفی از تصویر با دورانی با زاویه‌ای تصادفی بین صفر تا نود درجه و با آینه عمودی و افقی تصادفی قرار گیرد. پیش از قرارگیری تصویر بازتاب روی تصویر اولیه، تصویر با فیلتر گاوی با واریانس صفر که معادل فیلتر تارکننده یکنواخت و تمام یک است فیلتر شده و شدت رنگ آن در عددی تصادفی که همان ضرب (I-w) موجود در رابطه (۱-۱) است و بین ۰,۰ و ۰,۳ ضرب می‌شود، سپس قسمت‌هایی از تصویر اصلی که قرار است با بازتاب جمع شود، در مکمل این عدد یعنی w ضرب می‌شود و در نهایت تصویر با بازتاب آن جمع می‌شود و تصویر حاوی بازتاب شکل می‌گیرد. دو نمونه از مجموعه داده تولید شده بر روی تصویر نمونه موجود در شکل (۳-۳) را در شکل‌های (۴-۳) و (۵-۳) می‌بینیم. همچنین دو نمونه از پنج تصویر استفاده شده به عنوان بازتاب را در شکل‌های (۶-۳) و (۷-۳) می‌بینیم.



شکل (۴-۳): نمونه اول تصویر ساخته شده از مجموعه داده CamVid



شکل (۳-۵): نمونه دوم تصویر ساخته شده از مجموعه داده CamVid



شکل (۳-۶): نمونه اول دوم تصویر بازتاب

۳-۱-۴-مجموعه داده حقیقی همراه با بازتاب طبیعی

شرکت ره بین صنعت نصیر در حوزه خودرو های خودران و ابزار های کمکی برای رانندگان فعالیت و می کند و یکی از محصولات آن سیستمی شامل دوربین ارزان قیمت است که قسمت های مختلف جاده و نیز تابلو ها را تشخیص می دهد ولی در صورت شدت گرفتن بازتاب قدرت تشخیص آن افت می کند. این شرکت تعدادی از فیلم های ضبط شده حین رانندگی در سطح شهر تهران را در اختیار ما گذاشته تا از آن برای سنجش میزان کاربردی بودن روش های پیشنهادی خود استفاده نماییم. بدیهی است که این مجموعه داده، قابل استفاده بر روی روش های مبتنی بر یادگیری نظارت شده^۱ که روش های پیشنهادی ما نیز در همین

^۱ Supervised Learning

مجموعه قرار می‌گیرد نخواهد بود. بنابرین تنها استفاده ما از این مجموعه داده، این خواهد بود که ببینیم شبکه‌های آموزش داده شده، چه مقدار خوب می‌توانند بر روی داده واقعی عمل کنند.

۳-۵- تقسیم بندی داده‌های موجود

می‌دانیم در کاربردهای یادگیری ماشین و یادگیری عمیق، حداقل دو دسته داده باید داشته باشیم، دسته‌ای برای آموزش و دسته‌ای برای تست مدل آموزش داده شده، چرا که اینگونه می‌توان متوجه شد شبکه چهار بیش برآش^۱ یا کم برآش^۲ نشده باشد. اگر شبکه روی داده‌های آموزشی خوب عمل کند، متوجه می‌شویم که شبکه به اندازه کافی آموزش دیده و چهار کم برآش نشده، در غیر این صورت یا به اندازه کافی آموزش ندیده و یا اینکه شبکه پیچیدگی کمی جهت یادگیری عملیات مورد نظر دارد. اگر هم بر روی داده آموزش و هم تست به خوبی عمل کرد، به این معنی است که شبکه به خوبی آموزش دیده و از پس عملیاتی که به آن سپرده شده است به خوبی بر می‌آید، در صورتی که روی داده آموزش خوب عمل کند ولی روی داده تست خوب عمل نکند، به این معنی است که شبکه چهار بیش برآش شده و احتمالاً داده‌های آموزش دارای تنوع کم یا تعداد کم می‌باشد و یا اینکه شبکه به اندازه کافی پیچیده نیست و نتوانسته ویژگی‌های مورد نظر را یاد بگیرد.

برای جدا سازی دو مجموعه، از ۱۰۱۸ تصویر از مجموعه SIRR، تعداد ۸۹۶ تصویر جهت آموزش و مابقی جهت تست جدا سازی شدند، و از مجموعه داده‌های تولید شده ابتدا ۶۷۲ تصویر از ۷۰۱ تصویر برای آموزش جهت تست کنار گذاشته شده و سپس از روی آن‌ها داده تولید کردیم. در نهایت مجموعه داده‌های و مابقی جهت تست کنار گذاشته شده و سپس از روی آن‌ها داده تولید کردیم. در نهایت مجموعه داده‌های آموزش دارای ۱۱۶۴۸ تصویر و مجموعه داده‌های تست دارای ۵۸۶ تصویر می‌باشد. دلیل اینکه مجموعه تست نسبت به آموزش نسبت تقریبی ۱ به ۲۰ دارد و حدود ۵ درصد داده‌ها را تشکیل می‌دهد، این است که نسبتاً حجم داده‌ها کم بوده است و علاوه بر آن نتیجه نهایی روی مجموعه تصاویر حقیقی قرار است تست

^۱ Overfitting

^۲ Underfitting

شود. از طرف دیگر، معیار دقیقی در میزان حذف بازتاب نداریم و عملاً باید به صورت چشمی نتایج را بررسی کنیم.

۲-۳- روشهای پیاده سازی شده

جهت حذف بازتاب از تصویر، دو روش نسبتاً مشابه ارائه شده که هر دوی آن‌ها مبتنی بر شبکه U-Net هستند. و هر دوی آن‌ها یک تفاوت اصلی با شبکه U-Net اصلی دارد و آن افزایش یافتن تعداد فیلتر‌های آخرین لایه از ۱ به ۳ می‌باشد، چرا که در U-Net اولیه هدف بخش بندی تصاویر پزشکی، به تصویری سیاه سفید بوده است و نیاز به وجود سه دسته خروجی در آخرین لایه نبوده است، ولی در کاربرد ما، تصاویر رنگی هستند بنابرین نیاز داریم تعداد لایه‌ها را از ۱ به ۳ افزایش دهیم تا بیانگر ۳ کanal پیکسل‌ها در تصویر باشد. برای پیاده سازی از فریمورک تنسورفلو ۲ بر روی پایتون ۳.۷ استفاده شده است.

در هر دو پیاده سازی از میانگین نرم دوم تفاوت تصویر حقیقی و خروجی به عنوان تابع هزینه استفاده شده است، در واقع اگر n را تعداد کل تصاویر و T' را خروجی شبکه و تصویری که شبکه به عنوان تصویری که بازتاب آن کاهش یافته در نظر بگیریم و T تصویر حقیقی عاری از بازتاب باشد، تابع هزینه را می‌توان با رابطه $(1-۳)$ نشان داد.

$$\frac{\sum_i^n (T'_i - T_i)^2}{n} : (1-3)$$

۲-۱- روشن اول

در این روش، شبکه U-Net اولیه با اعمال تغییری که پیشتر ذکر شد، بدون هیچ تغییر دیگری، توسط مجموعه داده آماده شده، به مدت حدودی ۱۶ ساعت و به تعداد epoch ۱۲۰ بر روی پلتفرم گوگل کولب^۱ و

^۱ Colab

تحت پردازنده گرافیکی آموزش دیده است. در نهایت، تابع هزینه این شبکه هنگام آموزش به مقدار حدودی $10^{-4} \times 1/1$ رسید.

۳-۲-۲-۳ روش دوم

این روش نیز مشابه روش اول است، با این تفاوت که تمامی فیلترهای پیچشی، از فیلتر با اندازه ۳ در ۳، با فیلتر با اندازه ۵ در ۵ جایگزین شده اند، دلیل آن اثبات کارآمدی فیلترهای ۵ در ۵ در مقاله [۱۵] می‌باشد. روال آموزش مشابه روش اول است با این تفاوت که مدت زمان ناچیزی در حدود ۱ الی ۲ ساعت بیشتر صرف یادگیری شده است. در نهایت تابع هزینه این شبکه هنگام آموزش به مقدار $10^{-4} \times 1/7$ رسید.

۳-۲-۳ پیش‌پردازش داده‌های ورودی

قبل از دادن تصاویر به عنوان ورودی به شبکه، مقادیر هر سه کanal هر پیکسل در تصاویر بر ۲۵۵ تقسیم شده و بین صفر و یک قرار می‌گیرد، این باعث می‌شود نیازی به تغییر لایه آخر شبکه از سیگموید به ReLU یا تابعی خطی که خروجی بیشتر از ۱ نیز بتواند بدهد نباشد و شبکه زودتر همگرا شود.

۴-۲-۳ آموزش

به دلیل کمبود امکانات، مجبور به پیاده‌سازی و آموزش روی پلتفرم رایگان کولب شدیم که به ازای هر ایمیل حدود ۴ الی ۸ ساعت کارت گرافیک در اختیار قرار میدهد و پس از آن تا مدت حداقل ۱۲ ساعت دیگر نمی‌توان از آن استفاده کرد. به همین دلیل هنگام آموزش هر epoch ۱۰، مقادیر شبکه ذخیره می‌شوند تا در صورت اتمام زمان، شبکه بر روی ایمیل دیگری ادامه آموزش خود را ببیند. واضح است که شکسته شدن آموزش، می‌تواند در نتیجه نهایی تاثیر بگذارد. از طرف دیگر به دلیل استفاده از کولب و اینکه در هر بار استفاده کارت گرافیک متفاوتی را در اختیار می‌گذارد، محاسبه زمان و در نظر داشتن آن کار سخت و در عین حال بیهوده ای محسوب می‌شود. برای آموزش شبکه، تصاویر حاوی بازتاب به شبکه داده شدند و هدف و خروجی شبکه، تصویر عاری از بازتاب تصویر داده شده بودند. بنابرین در این روش بازتاب به صورت مجزا بدست نمی‌آید.

فصل ۴- ارزیابی، نتیجه گیری و پیشنهاد ها

۱-۴- ارزیابی

در این قسمت به بررسی دقت و نتایج بدست آمده می پردازیم و نتایج دو روش پیاده سازی شده را بررسی می کنیم. در ابتدا عملکرد دو روش را روی داده های آموزش، سپس روی داده های تست و در نهایت روی داده های واقعی شرکت ره بین نصیر می بینیم.

۱-۱- دقت

پس از پایان آموزش هر دو مدل را روی داده های تست ارزیابی^۱ کردیم، مقدار تابع هزینه و دقت شبکه ها در جدول (۱-۴) آمده است. در اولین بررسی، متوجه می شویم که اولاً هر دو شبکه به خوبی یاد گرفته اند و دوم اینکه در حال حاضر شبکه ها دچار بیش یا کم برازش نشده اند، چون تابع هزینه آن ها مقداری کم داشته و مقدار تست و آموزش آن به هم نزدیک است. غیر از این به طور کلی متوجه می شویم که شبکه دوم عملکرد نسبتاً ضعیف تری به طور کلی نسبت به شبکه اول دارد، هر چند چندان تفاوتی نمی کند.

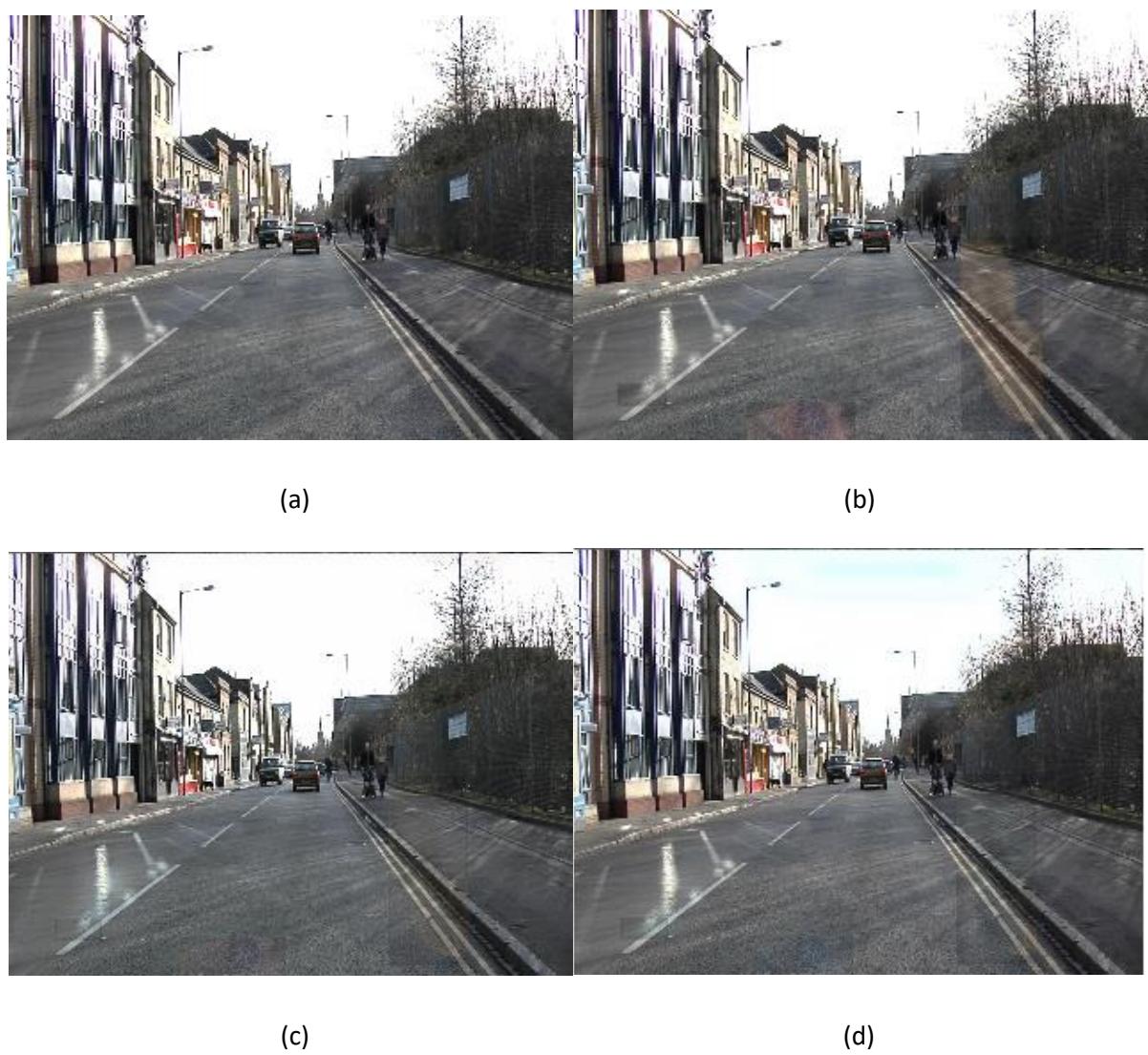
جدول ۱-۴- نتایج ارزیابی دو شبکه بر روی داده تست

هزینه	دقت	مدل
$3/6 \times 10^{-4}$	۷۶,۲	شبکه اول
$4/9 \times 10^{-4}$	۷۱,۳	شبکه دوم

^۱ Evaluation

۲-۱-۴ داده آموزش

در این قسمت مقایسه عملکرد دو شبکه را به شکل تصویر از مجموعه داده CamVid استفاده شده برای آموزش محدود می کنیم. مجموعه داده SIRR و عملکرد شبکه روی آن را بررسی نمی کنیم، چرا که بخش کوچکی از کل مجموعه داده و کمتر از ۱۰ درصد آن را تشکیل داده و طبعاً عملکرد چندان مناسبی را شاهد نخواهیم بود، همچنین هدف نهایی شباهت چندانی به محیط SIRR ندارد. اما اینکه در آموزش، از مجموعه داده SIRR استفاده شده بی تاثیر نبوده، چرا که تنوع بازتاب های آن بیشتر بوده و به یادگیری بهتر شبکه کمک نموده است. شکل (۱-۴) نتیجه اجرای شبکه ها بر روی یک نمونه داده تست را نشان می دهد.



شکل (۱-۴): نمونه تصویر آموزش و عملکرد شبکه ها روی آن – (a) تصویر اولیه، (b) تصویر حاوی بازتاب، (c) نتیجه شبکه دوم و (d) نتیجه شبکه اول می باشد.

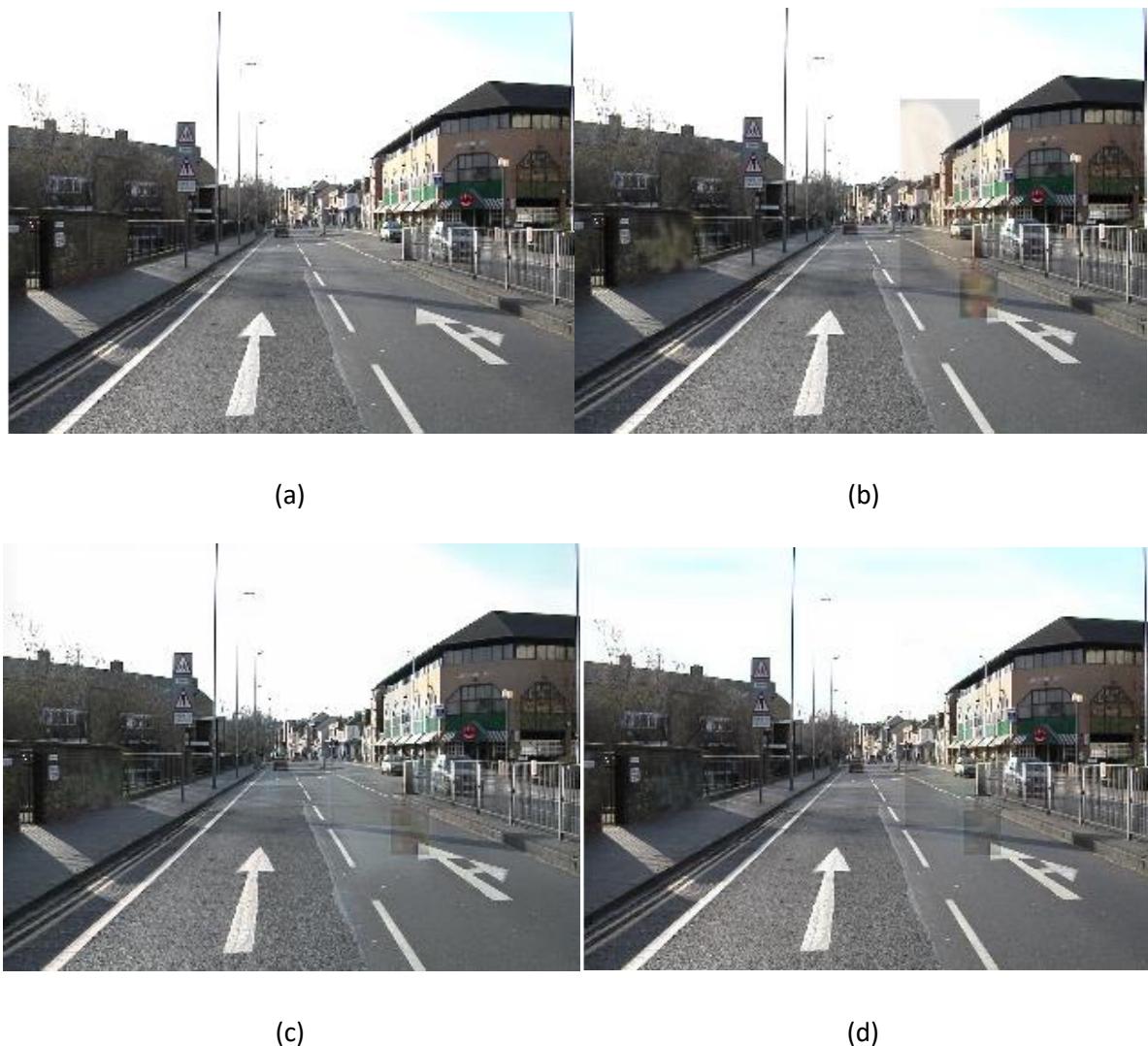
همانطور که از روی مقدار تابع هزینه بدست آمده دو شبکه انتظار می رود، شبکه اول بهتر توانسته بازتاب را حذف کند و در عین حال در هیچ قسمتی از تصویر شبکه دوم بهتر عمل نکرده است، اما به طور کلی تفاوت عملکرد دو شبکه بر روی داده آموزش، بسیار ناچیز است و در واقع شکل (۴-۱) نمونه ای از داده های آموزش است که تفاوت عملکرد دو شبکه تا حدی قابل مشاهده است.

۴-۳-داده تست

در ادامه نمونه ای از داده های تست از مجموعه داده CamVid را بررسی می کنیم. در شکل های (۴-۲) و (۴-۳) نتایج اجرای دو نمونه داده تست بر روی شبکه ها آمده است.



شکل (۴-۲): نمونه تصویر تست و عملکرد شبکه ها روی آن - (a) تصویر اولیه، (b) تصویر حاوی بازتاب، (c) نتیجه شبکه دوم و (d) نتیجه شبکه اول می باشد.



شکل (۴-۳): نمونه تصویر تست و عملکرد شبکه‌ها روی آن - تصویر (a) تصویر اولیه، تصویر (b) تصویر حاوی بازتاب، تصویر (c) نتیجه شبکه دوم و (d) نتیجه شبکه اول می‌باشد.

بر خلاف انتظار، همانطور که در نتایج دیده می‌شود، در قسمت‌هایی از تصویر، به خصوص قسمت‌هایی که بازتاب، بخش بزرگتری از تصویر را پوشانده، شبکه دوم عملکرد بهتری داشته و در قسمت‌هایی که جزئیات کوچک هستند، شبکه اول عملکرد بهتری داشته است، بنابرین نمی‌توان به طور قطع، یک معماری را بر دیگری ارجحیت داد.

۴-۱-۴-داده حقیقی با بازتاب طبیعی

در این قسمت دو نمونه تصویر واقعی و عملکرد شبکه روی آن ها را بررسی می کنیم تا ببینیم مجموعه داده ای که با استفاده از CamVid به صورت مصنوعی ساختیم چقدر به دنیای واقعی نزدیک است. در شکل های (۴-۴) و (۴-۵) به ترتیب از چپ به راست تصویر اولیه و خروجی شبکه اول و دوم را می بینیم.



(a)

(b)

(c)

شکل (۴-۴): نمونه تصویر واقعی با بازتاب طبیعی - جاده خالی - تصویر (a) تصویر اولیه، (b) خروجی شبکه اول و (c) خروجی شبکه دوم می باشد.



(a)

(b)

(c)

شکل (۴-۵): نمونه تصویر واقعی با بازتاب طبیعی - جاده شلوغ - تصویر (a) تصویر اولیه، (b) خروجی شبکه اول و (c) خروجی شبکه دوم می باشد.

همانطور که در عکس ها دیده می شود، در تصاویر میانی که مربوط به شبکه اول است، لبه های نازک و گوشه ها به طور نسبی بهتر محو شده اند، ولی در قسمت هایی که بازتاب شی بزرگتری داشته ایم و شی به صورت یکنواخت و بدون لبه بوده است، توسط شبکه دوم بهتر محو شده است. البته که دلیل آن این است که فیلتر های پیچشی شبکه اول، به دلیل کوچکتر بودن ویژگی های ریزتر و فیلتر های شبکه دوم به دلیل

بزرگتر بودن، ویژگی‌های بزرگتری را استخراج کرده‌اند، و قسمت‌هایی که بازتاب بسیار بزرگی داشته‌اند در هیچ کدام، چندان محو نشده‌اند.

۴-۲- نتیجه گیری

همانطور که در بررسی‌ها مشاهده شد، به خوبی دریافته شد که شبکه اول بازتاب‌های با جزئیات زیاد و شبکه دوم بازتاب‌های بزرگتر ولی با جزئیات کمتر را می‌تواند محو کند. بنابرین می‌توان نتیجه گرفت عبارتی شبکه U-Net معماری نسبتاً مناسبی برای حذف بازتاب اشیاء از تصاویر است، ولی لزوماً یکسان بودن اندازه فیلتر‌های پیچش در همه لایه‌ها، ایده مناسبی نیست.

از طرف دیگر، مدت زمان یادگیری شبکه نسبت به اندازه مجموعه داده موجود چندان زیاد نبوده و ممکن است با آموزش بیشتر و با مدت طولانی‌تر شبکه، کیفیت عملکرد شبکه‌ها بالاتر رود که در صورتی ممکن است که سیستم قدرتمندی در اختیار باشد و دشواری‌ها و محدودیت‌های محیط گوگل کولب را نداشته باشد.

نتیجه دیگری که از این پایان‌نامه می‌توان گرفت، این است که مجموعه داده CamVid و نیز ساخت داده مصنوعی همراه با بازتاب از روی آن، بسیار می‌تواند برای آموزش شبکه مفید بوده و در تصاویر حقیقی استفاده گردد.

۴-۳- پیشنهادها

بر اساس نتیجه گیری‌های انجام شده، در اولین مرحله پیشنهاد می‌شود، شبکه دیگری پیاده سازی و تحت آموزش قرار بگیرد که همان معماری U-Net را داشته باشد ولی اندازه فیلتر‌های پیچشی یکسان نباشد، بلکه به این صورت باشد که در دو لایه اول رمزگذار و دو لایه آخر رمزگذار از فیلتر ۳ در ۳ جهت در آوردن ویژگی‌ها ریز و کوچک استفاده شود، سپس در دو لایه دوم رمزگذار و سوم و نیز دو لایه یکی به آخر و دولایه دو تا به آخر رمزگشایی از فیلتر‌های ۵ در ۵ استفاده شود و در مابقی لایه‌ها که دولایه آخر قسمت رمزگذار

و دو لایه اول قسمت رمزگشا و نیز دو لایه پل هستند از فیلتر های ۹ در ۹ جهت در آوردن بزرگترین بازتاب ها استفاده شود. علاوه بر این پیشنهاد می شود شبکه بر روی سیستمی پایدار تر نسبت به کولب و برای مدت طولانی تری تحت آموزش قرار بگیرد تا از عملکرد صد در صدی شبکه مطمئن شویم.

علاوه بر پیشنهاد بالا، از آنجایی که از شبکه U-Net در کاربردهای مبتنی بر شبکه های مولد رقابتی همچون [۲۲] بسیار به کار رفته اند و نتایج خوبی داشته اند و از طرفی با کمبود داده جهت تست مواجه هستیم، شاید اگر در کنار معماری U-Net از شبکه دیگری استفاده شود که در کنار U-Net GAN همچون [۲۲] تشکیل یک شبکه عصبی مولد را بدهد، نتیجه نهایی بهتری گرفته شود.

پیشنهاد دیگری که می توان داد، بهبود تابع هزینه و استفاده از نرم اول به جای نرم دوم است، چرا که از [۲۲] می دانیم که نرم ۱ باعث می شود هنگام تولید تصویر، تصویر خروجی کمتر تار شود.

از آنجا که معمولا در خودرو های خودران معمولا با ویدیو سر و کار داریم، بنابرین میتوان از اطلاعات دنباله ای از تصاویر و ثبات نسبی بازتاب اشیاء در فریم های گوناگون ویدیو استفاده کرد. بدین جهت، شاید اگر از شبکه هایی که مبتنی بر دنباله ای از داده ها، همچون حافظه کوتاه-مدت ماندگار، استفاده کرد نتیجه نهایی به مراتب کیفیت بالاتری داشته باشد. حتی اگر از چنین شبکه ها و ابزار های مرتبط با تحلیل دنباله ای از داده استفاده نشود، و از یک شبکه تک تصویره استفاده شود، پیشنهاد می شود هر چند فریم، تصویر به شبکه داده شود، تفاوت تصویر اولیه و خروجی شبکه مقایسه شده و بازتاب را بدست آوریم، سپس در فریم های بعدی بازتاب را از تصویر کم کنیم و قسمت های کم شده را با ضرب در مقداری و یا سایر تکنیک ها، مجدد تقویت کنیم.

در نهایت، پیشنهاد می شود در صورت تصمیم به استفاده از شبکه های بررسی شده، دو شبکه را در طول یکدیگر به صورت آبشاری^۱ قرار داده و تصویر ابتدا به شبکه اول و سپس به شبکه دوم داده شود تا هر دو نوع

^۱ Cascade

حذف بازتاب را داشته باشیم، غیر از این می‌توان از هر تک شبکه چندین بار به صورت آبشاری استفاده کرد و تصویر را چندین بار به شبکه داده و خروجی را مجدد به ورودی دهیم.

پیوست ا- واژه‌نامه فارسی-انگلیسی

واژه فارسی	Equivalent English	واژه فارسی	Equivalent English
رمزنگاش	Decoder	ارتباطات همراه با پرسش	Skip connections
روشنایی	Brightness	ارزیابی	Evaluation
شار نوری	Optical Flow	الحاق	Concatenation
شبکه پیچشی عمیق	Deep Convolutional Network	انتشار روبه جلو	Forward Propagation
شبکه تماما پیچشی	Fully Convolutional Network	رمزنگار	Encoder
شبکه تماما متصل	Fully Connected Neural Network	ارمزنگار-رمزنگاش	Encoder-Decoder
مانندگار	شبکه حافظه کوتاه-مدت	آبشار	Cascade
باپاس	Long Short-Term Memory Network	آرایه	Array
بخش بندی تصویر	شبکه عصبی	آموخت	Train
بیش برآش	Generative Adversarial Network	باپاس	Bias
بینایی کامپیوتری	فیلتر	بخش بندی تصویر	Image Segmentation
پارامتر	Blurring Gaussian Filter	بیش برآش	Overfitting
پردازش تصویر	فیلتر تارکننده گاوسی	بینایی کامپیوتری	Computer Vision
پس انتشار	Underfitting	پارامتر	Parameter
پل	Colab	پردازش تصویر	Image Processing
پیچش	Support Vector Machine	پس انتشار	Back Propagation
پیکسل	Dataset	پل	Bridge
تبدیل تصویر به تصویر	مشکل گم شدن گرادیان ها	پیچش	Convolution
تبدیل تصویر به تصویر	Vanishing Gradients Problem	پیکسل	Pixel
پیچش ترانهاده	Neuron	تبدیل تصویر به تصویر	Image-to-Image Translation
تشخیص اشیاء	Weight	پیچش ترانهاده	Transposed Convolution
تکرار	Feature	تشخیص اشیاء	Object Detection
تنک	Spatial Features	تکرار	Iteration
			Sparse

واژه فارسی	Equivalent English
یادگیری عمیق	Deep Learning
یادگیری ماشین	Machine Learning
یادگیری ناظارت شده	Supervised Learning

پیوست ب - واژه‌نامه انگلیسی-فارسی

واژه فارسی	واژه انگلیسی
آرایه	Array
پس انتشار	Back Propagation
بایاس	Bias
فیلتر تارکننده گاوسی	Gaussian Filter
پل	Bridge
روشنایی	Brightness
آبشار	Cascade
کولب	Colab
بینایی کامپیوتری	Computer Vision
الحاق	Concatenation
پیچش	Convolution
مجموعه داده	Dataset
رمزنگشا	Decoder
شبکه پیچشی عمیق	Deep Convolutional Network
یادگیری عمیق	Deep Learning
رمزنگذار	Encoder
رمزنگذار-رمزنگشا	Encoder-Decoder
ارزیابی	Evaluation
ویژگی	Feature
فیلتر	Filter
انتشار روبه جلو	Forward Propagation
تنک	Sparse
ارتبطات همراه با پرش	Skip connections
پیکسل	Pixel
پارامتر	Parameter
بیش برآذش	Overfitting
شار نوری	Optical Flow
تشخیص اشیاء	Object Detection
نورون	Neuron
شبکه عصبی	Neural Network
یادگیری ماشین	Machine Learning
شبکه حافظه کوتاه-مدت ماندگار	Long Short-Term Memory Network
تکرار	Iteration
بردازش تصویر	Image Processing
بخش بندی تصویر	Image Segmentation
تبديل تصویر به تصویر	Image-to-Image Translation
شبکه مولد روابطی	Generative Adversarial Network
شبکه تمام پیچشی	Fully Convolutional Network
شبکه تمام متصل	Fully Connected Neural Network
آرایه	Equivalent English

واژه فارسی	Equivalent English
ویژگی‌های محلی	Spatial Features
یادگیری ناظارت شده	Supervised Learning
ماشین بردار پشتیبان	Support Vector Machine
آموزش	Train
پیچش ترانهاده	Transposed Convolution
کم برازش	Underfitting
مشکل گم شدن گرادیان‌ها	Vanishing Gradients Problem
وزن	Weight

فهرست مراجع ها

- [۱] A. Levin, A. Zomet, and Y. Weiss. Learning to perceive transparency from the statistics of natural scenes. In S. Becker, S. Thrun, and K. Obermayer, editors, Advances in Neural Information Processing Systems 15, 2002.
- [۲] A. Levin, A. Zomet, and Y. Weiss. Separating reflections from a single image using local features. 2004.
- [۳] A. Levin and Y. Weiss. User assisted separation of reflections from a single image using a sparsity prior. Computer VisionECCV 2004, pp. 602–613, 2004.
- [۴] Yang, Y.; Ma, W.; Zheng, Y.; Cai, J.F.; Xu, W. Fast single image reflection suppression via convex optimization. In CVPR, pp. 8141–8149, 2019.
- [۵] Y. Li and M. S. Brown. Single image layer separation using relative smoothness. In CVPR, pp. 2752–2759, 2014.
- [۶] B. Sarel and M. Irani. Separating transparent layers through layer information exchange. In Proc. of European Conference on Computer Vision, pp. 328–341, 2004.
- [۷] A. Agrawal, R. Raskar, S. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flashexposure sampling. ACM Trans. Graphics, 23(3):828–835, 2005.
- [۸] R. Szeliski, S. Avidan, and P. Anandan. Layer extraction from multiple images containing reflections and transparency. In CVPR, pp. 246–253, 2000.
- [۹] Y. Tsin, S. Kang, and R. Szeliski. Stereo matching with linear superposition of layers. IEEE TPAMI, 28(2): pp. 290–301, 2006.
- [۱۰] Yang, J., Li, H., Dai, Y., Tan, R.T.: Robust optical flow estimation of double-layer images under transparency or reflection. In CVPR. pp. 1410–1419, 2016.
- [۱۱] Ajay Nandoriya, Mohamed Elgharib, Changil Kim, Mohamed Hefeeda, and Wojciech Matusik. Video reflection removal through spatio-temporal optimization. In Proceedings of the IEEE International Conference on Computer Vision, pp. 2411–2419, 2017.
- [۱۲] Ahmed, Amgad et al. “User-assisted video reflection removal.” Proceedings of the 12th ACM Multimedia Systems Conference, 2021.
- [۱۳] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman. Reflection removal using ghosting cues. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3193–3201, 2015.

- [14] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David Wipf. A generic deep architecture for single image reflection removal and image smoothing. In Proceedings of the IEEE International Conference on Computer Vision, pp. 3238– 3247, 2017.
- [15] Zhixiang Chi, Xiaolin Wu, Xiao Shu, and Jinjin Gu. Single image reflection removal using deep encoder-decoder network. ArXiv:1802.00094, 2018.
- [16] Xuaner Zhang, Ren Ng, and Qifeng Chen. Single image reflection separation with perceptual losses. In CVPR, pp. 4786–4794, 2018.
- [17] Chao Li, Yixiao Yang, Kun He, Stephen Lin, and John E Hopcroft. Single image reflection removal through cascaded refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp/ 3565– 3574, 2020.
- [18] "Single-image-reflection-removal-dataset,"
<https://www.kaggle.com/siboooo/singleimagereflectionremovaldataset>.
- [19] "CamVid Database," <http://web4.cs.ucl.ac.uk/staff/g.brostow/MotionSegRecData/>
- [20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, pp. 234– 241, Springer, 2015.
- [21] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pages 448--456, 2015.
- [22] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967-5976, 2017.
- [23] "Artificial Neural Network," (10 July 2022) Wikipedia.
https://en.wikipedia.org/wiki/Artificial_neural_network
- [24] K E, S. (n.d.). Convolution Neural Networks. Developers Breach. Retrieved July 10th 2022, from <https://developersbreach.com/convolution-neural-network-deep-learning/>

Abstract

In recent years, we have observed many advancements in the fields of Computer Vision and Image Processing, and consequently, self-driving vehicles. One of the main responsibilities of self-driving vehicles is to be able to detect objects correctly and quickly, which has become very easy to achieve due to the emergence of neural networks. However, there is a fundamental problem regarding the cameras that are being used in such vehicles, which is being placed behind the windshield. This causes unwanted reflections of in-vehicle objects on the windshield, which heavily degrades the quality of captured images, and consequently leads to poor performance of the object detection system, due to being trained with images captured in an ideal environment with no reflection or dirt. Thus, there is an inevitable need for a reflection removal system that removes the reflection or at least decreases its effect.

In this thesis, we use a dataset to create images that contain reflections, then we train two U-Net based neural networks with it and achieve an accuracy of over 71 percent in each one of them. At the end, we compare the two models and reach a conclusion, and some notes on how to improve the models and build a better model.

key words: Image Reflection Removal, Layer Separation, Image Enhancement, Video Reflection Removal, U-Net, Single Image Reflection Removal



K. N. Toosi University of Technology
Faculty of Computer Engineering

**A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Bachelor of Science (B.Sc.)
in Computer Engineering**

Reflection Removal from In-Vehicle Images

By:

Seyed Shayan Daneshvar

Supervisor:

Seyed Behrooz Nasihatkon

Advisor:

Babak Nasersharif

Summer 2022