



به نام خدا  
دانشگاه تهران  
دانشکده مهندسی برق و کامپیوتر



## درس شبکه‌های عصبی و یادگیری عمیق

### تمرین امتیازی

| نام دانشجو | پرسش           |
|------------|----------------|
| ثمین سلوکی | شماره دانشجویی |
|            |                |
|            |                |

## فهرست

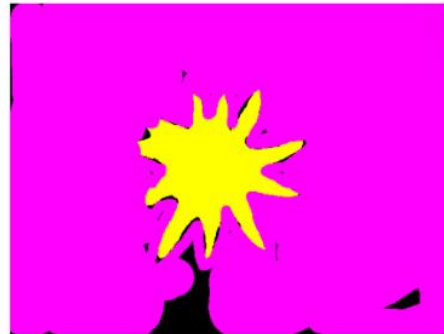
|    |   |
|----|---|
| ۲  | پرسش ۴) شبکه بخش بندی تصاویر - شبکه های UNet و Ta-UNet            |
| ۲  | ۴-۱) دادگان   |
| ۳  | مزایای داده های متوازن  |
| ۴  | تکنیک هایی برای مدیریت داده های نامتعادل                          |
| ۴  | تکنیک هایی برای مدیریت داده های نامتعادل در Semantic segmentation |
| ۵  | نرمالسازی   |
| ۹  | داده افزایی   |
| ۱۱ | ۴-۲) شبکه مورد استفاده  |
| ۱۱ | UNet  |
| ۱۳ | TA-UNet   |
| ۱۴ | ۴-۳) آموزش شبکه   |
| ۱۵ | ۴-۴) ارزیابی و تحلیل نتایج  |
| ۱۷ | مدل اصلی ۱) UNET  |
| ۱۸ | مدل اصلی ۲) TA-UNET   |

## شکل‌ها

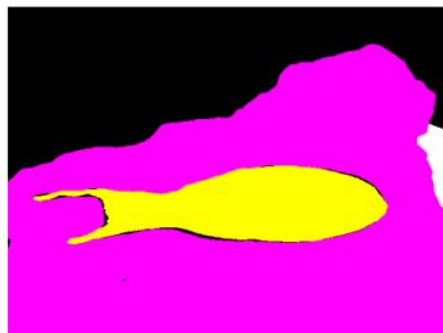
- شکل ۱- سمت چپ: تصویر RGB - سمت راست: ماسک مربوطه ..... ۲
- شکل ۲- سمت چپ: تصویر RGB - سمت راست: ماسک مربوطه ..... ۲
- شکل ۳- Z-score Normalization ..... ۶
- شکل ۴- Min-Max Normalization ..... ۶
- شکل ۵- Histogram Equalization ..... ۷
- شکل ۶- Adaptive Histogram Equalization (AHE) ..... ۷
- شکل ۷- CLAHE (Contrast Limited Adaptive Histogram Equalization) ..... ۸
- شکل ۸- Color Normalization (White Balance Correction) ..... ۸
- شکل ۹- Gamma Correction ..... ۹
- شکل ۱۰- داده افزایی ..... ۱۰
- شکل ۱۱- Unet ..... ۱۱
- شکل ۱۲- یادگیری مدل ..... ۱۴
- شکل ۱۳- UNET ..... ۱۷
- شکل ۱۴- TAUNET - ..... ۱۸

#### ۴-۱) دادگان

بخش آموزش و اعتبارسنجی دیتاست مورد نظر شامل ۱۵۲۵ عکس به همراه ماسک مربوطه است. دو نمونه از عکس و ماسک مربوطه را در زیر مشاهده می کنید:



شکل ۱- سمت چپ: تصویر RGB - سمت راست: ماسک مربوطه



شکل ۲- سمت چپ: تصویر RGB - سمت راست: ماسک مربوطه

در ادامه برای انجام آموزش و اعتبارسنجی، دادگان این دیتاست به نسبت ۹۰-۱۰ به ترتیب برای آموزش و اعتبارسنجی تقسیم شدند. در همین ابتدا ذکر می کنیم که ابعاد عکس ها  $360 \times 640$  بوده اما طبق مقاله ی پایه، ابعاد تصاویر به  $368 \times 640$  تغییر یافتند. کلاس ماسک ها برخلاف مقاله ی پایه ۲ نیست و طبق دیتاست استفاده شده برابر ۸ می باشد. در این پروژه سعی شد segmentation برای ۸ کلاس های دیتاست انجام شود.

در زمینه یادگیری ماشین و یادگیری عمیق، اطمینان از متعادل بودن داده های آموزش و ارزیابی بسیار مهم است. از دلایل و مزایای اصلی انتخاب دیتای آموزش متوازن و داده های ارزیابی به موارد زیر اشاره می کنیم:

جلوگیری از ایجاد بایس در مدل: وقتی مجموعه داده نامتعادل باشد، مدل می تواند به سمت کلاس اکثریت سوگیری پیدا کند و منجر به عملکرد ضعیف در کلاس اقلیت شود. به عنوان مثال، در یک مسئله طبقه بندی باینری با ۹۰ درصد نمونه ها در کلاس A و ۱۰ درصد در کلاس B، یک مدل ممکن است به سادگی کلاس A را همیشه برای دستیابی به دقت بالا پیش بینی کند، اما این گمراه کننده خواهد بود. داده های متوازن به مدل کمک می کند تا الگوها و ویژگی های اساسی هر کلاس را به طور موثرتر یاد بگیرد و به یک مدل کلی تر و قوی تر بسازد.

بهبود معیارهای عملکرد: در مجموعه داده های نامتعادل، دقت یک معیار عملکرد قابل اعتماد نیست زیرا یک مدل می تواند با نادیده گرفتن کلاس اقلیت به دقت بالایی دست یابد. داده های متوازن تضمین می کند که معیارهایی مانند دقت، یادآوری، امتیاز F1 و ROC-AUC بازتابی واقعی از عملکرد مدل ارائه می دهند.

افزایش پایداری مدل: مجموعه داده های نامتعادل می تواند منجر به برآزش بیش از حد در کلاس اکثریت شود. مجموعه داده های متوازن با اطمینان از اینکه مدل فقط کلاس اکثریت را یاد نمی گیرد، بلکه تمایز بین کلاس های مختلف را یاد می گیرد، به کاهش بیش برآزش کمک می کند. به عبارت دیگر مدلی که بر روی داده های متوازن آموزش دیده باشد، به خوبی به داده های دیده نشده تعمیم می یابد، زیرا یاد گرفته است که کلاس های مختلف را به همان اندازه به خوبی مدیریت کند.

### مزایای داده های متوازن

عملکرد پیش بینی بهبود یافته: هر کلاسی که به طور مساوی نمایش داده می شود به مدل اجازه می دهد تا ویژگی های همه کلاس ها را بیاموزد و توانایی آن را برای پیش بینی های دقیق در سراسر صفحه بهبود بخشد.

تطبیق پذیری: چنین مدل هایی با تغییرات و تغییرات داده ها سازگارتر هستند و آنها را برای کاربردهای دنیای واقعی قابل اعتمادتر می کند.

استفاده موثر از معیارهای عملکرد: مجموعه داده های متوازن تضمین می کند که معیارهای عملکرد مانند ماتریس سردرگمی، منحنی های ROC و سایر ابزارهای ارزیابی تصویری جامع و دقیق از عملکرد مدل ارائه دهند و با معیارهای قابل اعتماد، دانشمندان داده و ذینفعان می توانند تصمیمات آگاهانه تری در مورد استقرار مدل، بهبودها و عیب یابی بگیرند.

انصاف و ملاحظات اخلاقی و اجتناب از تبعیض: در مدل هایی که fairness بسیار مهم است، مانند مراقبت های بهداشتی، مالی، و عدالت کیفری، مجموعه داده های متوازن به جلوگیری از تعصب مدل ها علیه گروه های کم نمایندگی کمک می کند. اطمینان از داده های متوازن به توسعه سیستم های هوش مصنوعی اخلاقی کمک می کند که نتایج عادلانه ای را در بین جمعیت های مختلف و موارد استفاده ارائه دهد.

### تکنیک هایی برای مدیریت داده های نامتعادل

- نمونه گیری مجدد: یا با نمونه برداری بیش از حد از کلاس اقلیت (به عنوان مثال، SMOTE) یا کم نمونه برداری از طبقه اکثریت.
- Synthetic Data Generation: ایجاد نمونه های مصنوعی برای تقویت کلاس اقلیت.
- وزن بندی کلاس: اختصاص وزن های بالاتر به کلاس اقلیت در حین تمرین برای جریمه بیشتر مدل برای رده بندی های اشتباه.

### تکنیک هایی برای مدیریت داده های نامتعادل در Semantic segmentation

عدم تعادل در Semantic segmentation زمانی رخ می دهد که کلاس های خاصی (مثلاً پس زمینه) بر تصاویر تسلط دارند، در حالی که برخی دیگر (مثلاً اشیاء نادر) کمتر نمایش داده می شوند. در اینجا به چندین تکنیک برای پرداختن به داده های نامتعادل ارائه می کنیم:

- وزندهی مجدد کلاس:

تنظیم عملکرد Loss: اصلاح تابع Loss برای اختصاص وزن های بالاتر به کلاس های کمتر ارائه شده می تواند کمک کند. به عنوان مثال، با استفاده از cross-entropy که در آن هر کلاس دارای وزنی است که با فرکانس آن در مجموعه داده نسبت معکوس دارد.

استفاده از Focal Loss: این تابع به طور پویا افت آنتروپی متقابل را مقیاس می کند و تمرکز بیشتری را به نمونه هایی که طبقه بندی آنها دشوار است، که اغلب به کلاس های اقلیت تعلق دارند، می دهد.

- افزایش داده:

تولید داده مصنوعی: ایجاد نمونه های مصنوعی از کلاس های اقلیت می تواند به تعادل مجموعه داده کمک کند. تکنیک هایی مانند تقویت تصویر (چرخش، چرخش، مقیاس بندی) می توانند تعداد نمونه های کلاس اقلیت را افزایش دهند. راهکار دیگر، کپی کردن اشیاء کلاس اقلیت از یک قسمت از تصویر و چسباندن آنها در مکان های مختلف یا تصاویر مختلف برای افزایش مصنوعی فرکانس آنها است.

استراتژی های نمونه گیری مبتنی بر داده (نمونه گیری متوازن کلاس): اطمینان از اینکه هر مینی بچ در طول آموزش حاوی نمایش متعادلی از کلاس ها باشد. این می تواند شامل نمونه برداری انتخابی از پیکسل ها یا وصله های تصویر بر اساس توزیع کلاس باشد. همچنین می توان بر نمونه هایی که طبقه بندی سختی دارند (hard-to-classify examples)، که اغلب متعلق به طبقات اقلیت هستند، با شناسایی و اولویت بندی پویا آنها در طول آموزش این مسئله را حل کرد.

#### ▪ معماری شبکه پیشرفته:

مکانیسم های توجه: ترکیب مکانیسم های توجه برای تمرکز بر مناطق مهم تصویر که ممکن است شامل طبقات اقلیت باشد. (استفاده از معماری هایی که ویژگی ها را در مقیاس های متعدد ثبت می کنند تا اطمینان حاصل شود که اشیاء کوچک یا کلاس های اقلیت از دست نمی روند).

استفاده از تکنیک های پس از پردازش مثل CRF (Conditional Random Fields) یا شبکه های Segmentation Refinement

#### ▪ Loss متوازن طبقاتی:

Dice Loss: توابع زیان مبتنی بر ضریب تاس، که بر روی همپوشانی بین بخش های پیش بینی شده و واقعی تمرکز می کنند، به ویژه برای رسیدگی به وظایف تقسیم بندی نامتعادل موثر هستند.

Tversky Loss: تعمیم Dice Loss تاس که شامل پارامترهایی برای کنترل مبادله بین دقت و یادآوری است که به مدیریت بهتر عدم تعادل کلاس کمک می کند.

### نرمالسازی

برای انجام نرمالسازی داده ها می توان رویکرد های متفاوتی اتخاذ کرد که هر کدام بر دیگری برتری دارند. در ادامه به انواع نرمالسازی و نتیجه ی آن بر تصویر اشاره می کنیم.

#### Standardization (Z-score Normalization)

این روش داده ها را به میانگین 0 ( $\mu$ ) و انحراف معیار 1 ( $\sigma$ ) تبدیل می کند. به تثبیت واریانس و ایجاد ویژگی ها کمک می کند. اگر مجموعه داده دارای شرایط نوری متفاوتی باشد، استانداردسازی می تواند

موثر باشد. این به عادی سازی کنتراست و روشنایی در تصاویر مختلف کمک می کند.



شکل ۳- Z-score Normalization

### Min-Max Normalization

این روش داده ها را به یک محدوده ثابت، معمولاً  $[0, 1]$  یا  $[-1, 1]$  مقیاس می کند و زمانی مفید است که بخواهیم روابط بین مقادیر را حفظ کنیم. به ویژه زمانی که مقادیر پیکسل به دلیل شفافیت آب و نفوذ نور متفاوت در محدوده های مختلف قرار دارند، می تواند مفید باشد.



شکل ۴- Min-Max Normalization

### Histogram Equalization



این تکنیک کنتراست تصویر را با تغییر توزیع شدت تنظیم می کند. متداول ترین مقادیر شدت را پخش می کند و کنتراست کلی را بهبود می بخشد. تصاویر زیر آب اغلب از کنتراست کم و روشنایی متفاوت رنج می برند. یکسان سازی هیستوگرام می تواند دید اجسام را با بهبود کنتراست افزایش دهد.



شکل ۵ Histogram Equalization

#### Adaptive Histogram Equalization (AHE)

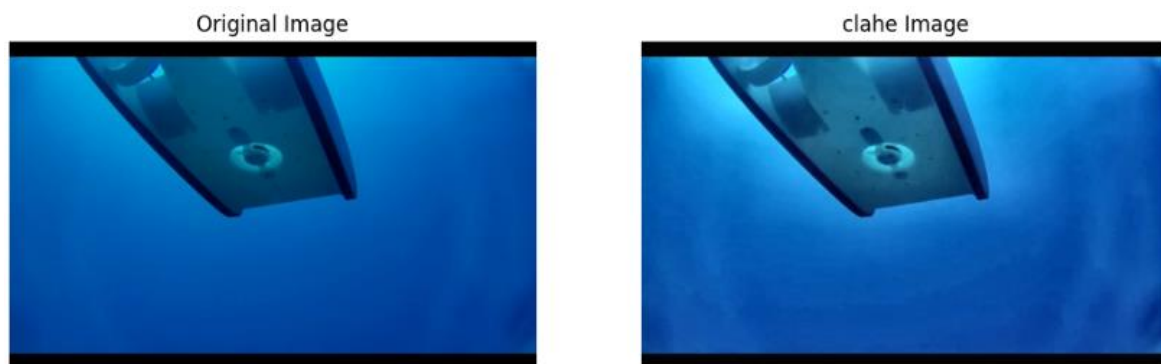
گونه ای از تساوی هیستوگرام، AHE کنتراست را با تقسیم تصویر به مناطق کوچکتر و اعمال یکسان سازی هیستوگرام برای هر ناحیه به طور جداگانه بهبود می بخشد. AHE به ویژه برای تصاویر با تغییرات محلی در روشنایی، که در محیط های زیر آب به دلیل پراکندگی و جذب نور رایج است، موثر است.



شکل ۶ Adaptive Histogram Equalization (AHE)

#### CLAHE (Contrast Limited Adaptive Histogram Equalization)

CLAHE شکلی از AHE است که با محدود کردن افزایش کنتراست از تقویت بیش از حد نویز جلوگیری می‌کند. CLAHE برای تصاویر زیر آب که نویز می‌تواند مشکل چالشی‌ای باشد مناسب است و کنتراست محلی را بدون تقویت بیش از حد نویز افزایش می‌دهد.



شکل ۷) CLAHE (Contrast Limited Adaptive Histogram Equalization)

#### Color Normalization (White Balance Correction)

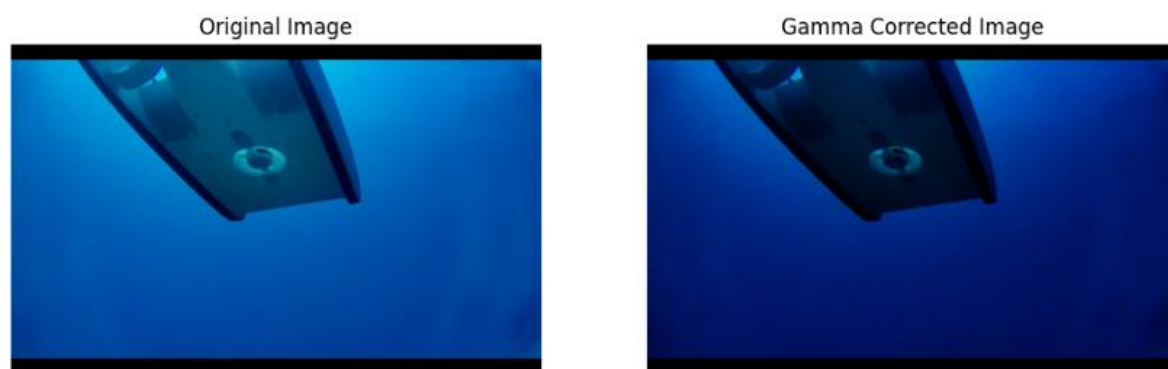
این روش تعادل رنگ یک تصویر را تنظیم می‌کند و اعوجاج رنگ به دلیل جذب و پراکندگی آب را تصحیح می‌کند. تصاویر زیر آب اغلب رنگ مایل به آبی یا سبز دارند. اصلاح تراز سفیدی به بازیابی رنگ‌های طبیعی کمک می‌کند و اشیاء را بیشتر متمایز می‌کند.



شکل ۸) Color Normalization (White Balance Correction)

#### Gamma Correction

تصحیح گاما روشنایی تصویر را تنظیم می‌کند و در عین حال سایه‌ها و هایلایت‌ها را حفظ می‌کند. برای اصلاح شرایط نور در تصاویر زیر آب مفید است و ویژگی‌ها را برجسته تر می‌کند.



شکل ۹ Gamma Correction

ما برای انجام این پروژه تصمیم گرفتیم روش `clahe`، `gamma_correction`، `white_balance` و `min max` و `scaling` را پیاده کنیم. این رویکرد ترکیبی از تکنیک‌های فوق، طبق تحقیقات قبلی اغلب بهترین نتایج را برای تصاویر زیر آب به همراه دارد. بدین ترتیب با تصحیح تراز سفیدی به اصلاح اعوجاج رنگ‌ها و طبیعی‌تر کردن رنگ‌ها می‌پردازیم و با تصحیح گاما برای تقویت تون‌های میانی و بهبود دید و با CLAHE به افزایش کنتراست محلی بدون تقویت بیش از حد نویز اقدام می‌کنیم.

تصحیح تراز سفیدی اعوجاج رنگ به دلیل جذب و پراکندگی آب را برطرف می‌کند. تصحیح گاما باعث بهبود دید رنگ‌های میانی می‌شود که برای تشخیص اجسام زیر آب بسیار مهم هستند. CLAHE کنتراست موضعی را افزایش می‌دهد و اشیاء را بدون افزایش بیش از حد نویز واضح‌تر نشان می‌دهد.

لازم به ذکر است که کدهای پیاده‌سازی تک تک نرمالسازی‌ها در فایل کد این پروژه آورده شده است.

### داده افزایی

همانطور که پیش از این ذکر شد، افزایش داده چندین مزیت کلیدی ارائه می‌دهد:

- ✓ افزایش داده‌ها به طور موثر اندازه مجموعه داده آموزشی را بدون نیاز به جمع‌آوری داده‌های جدید افزایش می‌دهد که می‌تواند پرهزینه و زمان‌بر باشد.
- ✓ تعمیم مدل را با قرار دادن مدل در معرض طیف گسترده‌تری از داده‌ها از طریق تبدیل‌ها، تقویت داده‌ها به تعمیم بهتر مدل به داده‌های دیده‌نشده کمک می‌کند و باعث کاهش بیش از حد برازش می‌شود.
- ✓ داده‌های افزوده تغییراتی را معرفی می‌کند (به عنوان مثال، چرخش، چرخش، تغییر رنگ)، مدل را نسبت به تغییرات و نویزهای دنیای واقعی قوی‌تر می‌کند.

✓ تقویت داده ها را می توان برای تولید نمونه های اضافی از کلاس های اقلیت استفاده کرد، که به تعادل مجموعه داده های نامتعادل و بهبود عملکرد در کلاس های کمتر ارائه شده کمک می کند.

ما برای انجام این پروژه، روش های زیر را اجرا کردیم:

```
# Data augmentation
data_gen_args = dict(rotation_range=0.2,
                      width_shift_range=0.05,
                      height_shift_range=0.05,
                      shear_range=0.05,
                      zoom_range=0.05,
                      horizontal_flip=True)
```

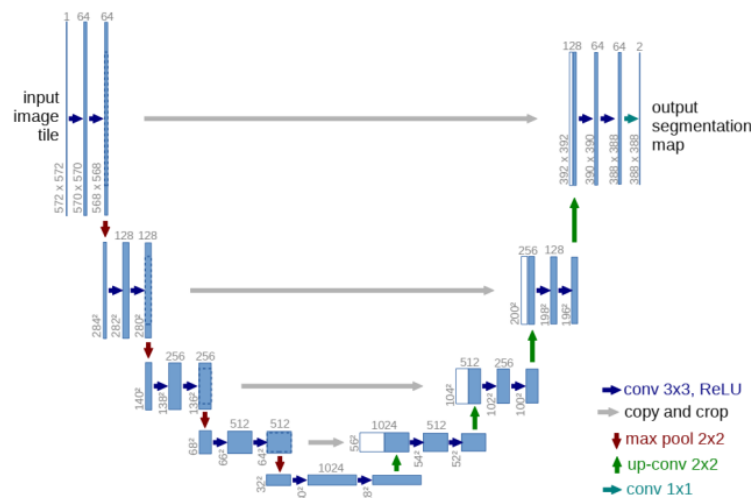
شکل ۱۰- داده افزایی

روش های داده افزایی بالا، همگی با توجه به این نکته انتخاب شده اند که داده ی خروجی این فرآیند حتما تابع توزیع اش تغییر نکرده باشد. تکنیک اول چرخش ۰.۲، فلیپ کردن، زوم کردن و ... همگی داده ای تولید می کنند که شباهت بسیاری با داده ها اولیه دارند. با توجه به این نکته که تصاویر مربوط به اشیا زیر آب هستند، آزادی عمل بیشتری برای اعمال این روش ها داریم. اما برای مثال اگر تصاویر MRI مغز بود باید دقت بیشتری به خرج می دادیم. اما اکنون که داده های ما عکس ماهی، ستاره دریایی، قایق و... می باشند، ما مطمئن هستیم که با روش های بالا، داده ی خروجی همچنان هم توزیع داده های ورودی است. (مثلا عکس یک غواص که ۰.۲ چرخیده شده همچنان یک غواص است و یا عکس ماهی ای که از عرض شیفست شده همچنان یک ماهی است!)

در ادامه ی گزارش، در بخش آموزش شبکه، به تاثیر استفاده از روش های داده افزایی بر عملکرد مدل بیشتر خواهیم پرداخت.

## شبکه مورد استفاده (۲-۴)

### UNet



شکل ۱۱- Unet

معماری U-Net شامل یک مسیر encoding و در ادامه یک مسیر decoding است. نام گذاری این شبکه به حرف U هم از همین فرایند کاهشی و سپس افزایشی الهام گرفته است.

در بخش encoding، متشکل از لایه های کانولوشنی به دنبال لایه های max-pooling می باشد. این لایه ها به تدریج ابعاد فضایی را پایین می آورند و برای جبران این کاهش سایز داده ی ورودی، تعداد کانال های ویژگی را دو برابر می کند. هر کانولوشن از لایه های کانولوشنال (اغلب با نرمالسازی دسته های

و

فعالسازی ReLU) تشکیل شده است که به دنبال آن 22 max-pooling با stride=2 برای کاهش ابعاد spatial انجام میشود. جزئیات معماری شبکه به صورت کامل در عکس بالا شرح داده شده است.

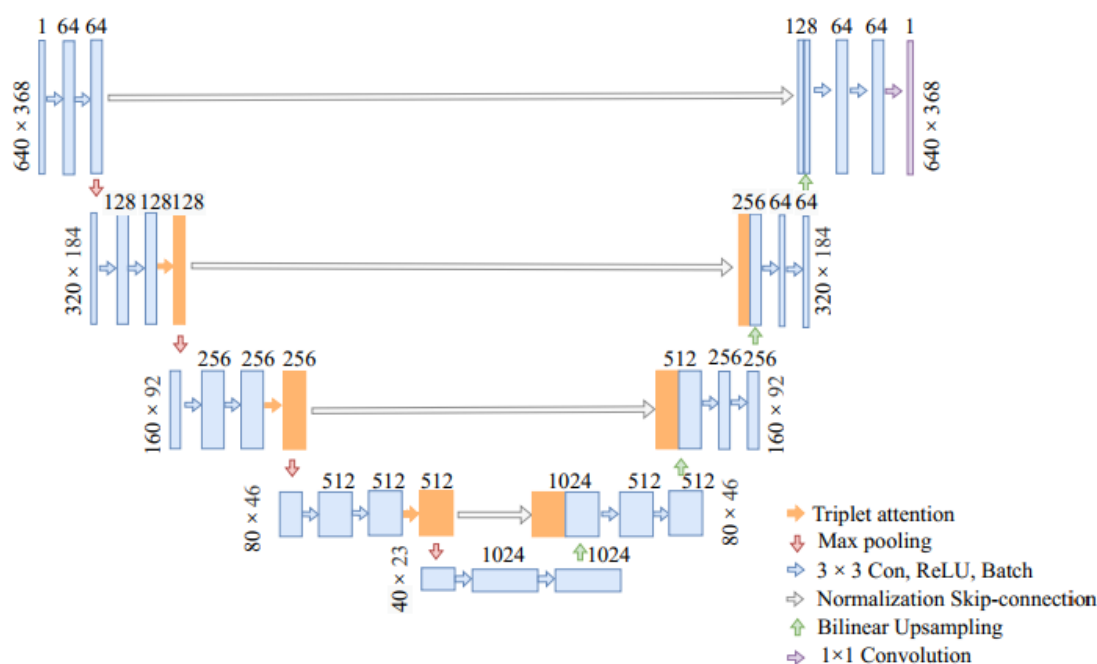
به منظور توضیح بیشتر معماری عکس شبکه می توان گفت که پس از ورود عکس ورودی و خروج از max-pooling اول، اندازه ی عکس از  $512 \times 512 \times 3$  به  $256 \times 256 \times 3$  و در قدم بعد پس از انجام level با هاینر پارامترهایی که پیش از این ذکر شد، ابعاد به  $128 \times 128 \times 3$  رسیده و وارد مرحله ی دوم می شود. در این مرحله هم پس از اعمال دو لایه کانولوشنی با کرنل  $3 \times 3$  و تابع فعالساز ReLU، ابعاد عکس به  $64 \times 64 \times 3$  می رسد. شاهد این هستیم که در encoding تعداد کانال های عکس به منظور جبران

کاهش رزولوشن، دو برابر می شود. (۶۴ به ۱۲۸ به ۵۱۲ به ۱۰۲۴). به طور خلاصه، encoding باعث کاهش رزولوشن فضایی اما تقویت چیهستی feature ها می شود.

بخش decoding، شامل لایه های upsampling و به دنبال آن لایه های کانولوشن است. این لایه ها feature map را به اندازه ورودی اولیه تبدیل میکنند و در عین حال تعداد کانالهای ویژگی را کاهش. در بخش encoding، متشکل از لایه های کانولوشنی به دنبال لایه های max-pooling می باشد. این لایه ها به تدریج ابعاد فضایی را پایین می آورند و برای جبران این کاهش سایز داده ی ورودی، تعداد کانال های ویژگی را دو برابر می کند. هر کانولوشن از لایه های کانولوشنال (اغلب با نرمالسازی دسته های و فعالسازی ReLU) تشکیل شده است که به دنبال آن  $2 \times 2$  max-pooling با  $\text{stride}=2$  برای کاهش ابعاد spatial انجام میشود. جزییات معماری شبکه به صورت کامل در عکس بالا شرح داده شده است. به منظور توضیح بیشتر معماری عکس شبکه می توان گفت که پس از ورود عکس ورودی و خروج از level اول، اندازه ی عکس از  $512 \times 512 \times 3$  به  $256 \times 256 \times 3$  و در قدم بعد پس از انجام max-pooling با هایپرپارامترهایی که پیش از این ذکر شد، ابعاد به  $128 \times 128 \times 3$  رسیده و وارد مرحله ی دوم می شود. در این مرحله هم پس از اعمال دو لایه کانولوشنی با کرنل  $3 \times 3$  و تابع فعالساز ReLU، ابعاد عکس به  $64 \times 64 \times 3$  می رسد. شاهد این هستیم که در encoding تعداد کانال های عکس به منظور جبران کاهش رزولوشن، دو برابر می شود. (۶۴ به ۱۲۸ به ۵۱۲ به ۱۰۲۴). به طور خلاصه، encoding باعث کاهش رزولوشن فضایی اما تقویت چیهستی feature ها می شود. بخش decoding، شامل لایه های upsampling و به دنبال آن لایه های کانولوشن است. این لایه ها feature map را به اندازه ورودی اولیه تبدیل میکنند و در عین حال تعداد کانالهای ویژگی را کاهش میدهند. این Upsampling می تواند با تکنیک هایی چون bilinear یا transposed convolutions .. انجام شود

بین هر مرحله ی downsampling و upsampling، از Skip connection ها استفاده می شود که feature map را از encoder به decoder به هم متصل میکنند و اطلاعات فضایی دقیقی را در طول upsampling ارائه می دهد. Skip connection نوآوری اصلی این مدل است که به شبکه اجازه میدهد اطلاعات فضایی دقیق را حفظ کند و همچنین از vanishing gradient جلوگیری می کند و فرآیند آموزش را بهبود میدهد. لایه نهایی U-Net معمولاً از یک لایه کانولوشن  $1 \times 1$  تشکیل شده است که به دنبال آن یک تابع فعال ساز (به عنوان مثال، sigmoid برای تقسیم بندی باینری یا softmax برای تقسیم بندی چند کلاسه) استفاده می شود. خروجی U-Net یک ماسک طبقه بندی (segmentation mask) با پیش بینی های پیکسلی است که در آن هر پیکسل نشان دهنده کلاس یا برچسب پیش بینی است.

## TA-UNet



TA-UNet یک معماری پیشرفته است که برای وظایف بخش‌بندی جاده در حوزه رانندگی خودمختار طراحی شده است. این معماری بر اساس ساختار U-Net ساخته شده است، که به دلیل اثربخشی آن در تقسیم‌بندی معنایی، به ویژه در تصویربرداری پزشکی شهرت دارد و اکنون برای نیازهای بصری پیچیده صحنه‌های جاده سازگار شده است. این معماری شامل یک چارچوب رمزگذار-رمزگشا با اتصالات پرش است که انتشار جزئیات فضایی سطح پایین را از رمزگذار به مراحل رمزگشا تسهیل می‌کند. این راه‌اندازی به مدل اجازه می‌دهد تا از زمینه محلی و جهانی استفاده کند و در عین حال اطلاعات دقیق را برای تقسیم‌بندی دقیق حفظ کند.

نوآوری TA-UNet، ادغام مکانیسم توجه سه‌گانه در مرحله رمزگذاری آن است. برخلاف مکانیسم‌های توجه مرسوم که اغلب فقط بر روی ابعاد کانال یا ابعاد فضایی جداگانه تمرکز می‌کنند، توجه سه‌گانه با گرفتن وابستگی‌ها در سه بعد به طور همزمان: کانال، ارتفاع و عرض، استخراج ویژگی را افزایش می‌دهد. این رویکرد در بخش‌بندی جاده‌ها، جایی که درک روابط فضایی بین بخش‌های مختلف صحنه (به عنوان مثال، علامت‌گذاری جاده، موانع) برای ناوبری دقیق ضروری است، بسیار مهم است.

ماژول توجه سه‌گانه از طریق سه شاخه موازی عمل می‌کند، که هر یک به ثبت تعاملات در ابعاد خاص تانسور ورودی اختصاص دارد. این شاخه‌ها از لایه‌های کانولوشنی و به دنبال آن توابع عادی سازی

و فعال سازی برای محاسبه وزن توجه استفاده می کنند. خروجی های این شاخه ها برای تولید نقشه های ویژگی تصفیه شده که اطلاعات مکانی دقیق را در بر می گیرد و به طور مؤثر مناطق مورد علاقه، مانند سطوح جاده های قابل رانندگی را در میان درهم رفتگی بصری پیچیده برجسته می کند، جمع می کنند. لازم به ذکر است که ماژول توجه سه گانه پس از لایه های کانولوشن اولیه در مرحله رمزگذاری معماری U-Net قرار می گیرد. این ادغام به حفظ جزئیات مهم فضایی و افزایش نمایش ویژگی ها کمک می کند.

علاوه بر این، TA-Unet به طور استراتژیک عملیات upsampling و convolutional را در مسیر رمزگشای خود برای بازسازی نقشه های تقسیم بندی با وضوح بالا از نمایش های ویژگی غنی شده تولید شده توسط رمزگذار، ترکیب می کند. این فرآیند تضمین می کند که مدل نه تنها مناطق جاده ای را شناسایی می کند، بلکه زمینه فضایی حیاتی برای درک طرح و تداوم شبکه راه را نیز حفظ می کند. با افزایش استخراج ویژگی از طریق توجه سه گانه و حفظ تقارن معماری با اتصالات پرش، TA-Unet عملکرد برتر را در معیارهای تقسیم بندی جاده ها نشان می دهد و به دقت و استحکام بالا برای کاربردهای دنیای واقعی در سیستم های رانندگی خودمختار دست می یابد.

#### ۳-۴ آموزش شبکه

ما پس از فراخوانی دیتا، نرمال سازی داده، و اجرای داده افزایی با تکنیک های متعدد و نهایتاً تقسیم داده به دو گروه اعتبارسنجی و آموزش به آموزش شبکه پرداختیم. در این پروژه دو مدل U-Net و TA-Unet آموزش دیدند. از آوردن کد دو مدل بدلیل طولانی بودن در این گزارش پرهیز کردیم.

در مقاله ی پایه ذکر شده است که یکی از روش های غلبه بر داده ی غیر متوازن، استفاده از توابع هزینه ی ترکیبی است. به همین دلیل ما در این پروژه دو تابع هزینه ی Cross entropy و Dice Loss را ترکیب کردیم.

```
checkpoint = ModelCheckpoint('best_model.h5', monitor='val_loss', save_best_only=True, mode='min')
reduce_lr = ReduceLROnPlateau(monitor='val_loss', factor=0.2, patience=3, min_lr=0.0005, mode='min')
early_stopping = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True, mode='min')

input_shape = (*img_size, 3)
num_classes = 8
model = UNET(input_shape, num_classes)
model.compile(optimizer='adam',
              loss=combined_loss,
              metrics=['accuracy', mean_iou])

history3 = model.fit(train_generator,
                    validation_data=val_generator,
                    epochs=20,
                    callbacks=[checkpoint, reduce_lr, early_stopping])
```

شکل ۱۲- یادگیری مدل



در فرآیند آموزش مدل، از `callback`، کاهش نرخ آموزشی و `earlystopping` استفاده کردیم و مدل را در ۱۵ و ۲۰ اپاک آموزش دادیم.

در `ModelCheckpoint Callback` این طور تعریف شده است که هر زمان که هزینه اعتبارسنجی (`val_loss`) کاهش یابد، این بازخوانی وزن های مدل را در یک فایل ذخیره می کند (`best_model.h5`) و تضمین می کند که بهترین مدل را بر اساس از دست دادن اعتبارسنجی حفظ می کنیم.

`ReduceLROnPlateau` فراخوانی نرخ یادگیری (`lr`) را زمانی که متریک (`val_loss`) در این مورد بهبود نمی یابد، کاهش می دهد. اگر بهبودی در از دست دادن اعتبار پس از ۳ دوره مشاهده نشد، با کاهش نرخ یادگیری به میزان ۰.۲ به همگرایی سریع تر مدل کمک می کند.

با `EarlyStopping` اگر از هزینه اعتبارسنجی (`val_loss`) پس از تعداد معینی از دوره ها بهبود نیابد، این فراخوانی فرآیند آموزش را زود متوقف می کند. «`restore_best_weights=True`» تضمین می کند که وزن های مدل به وزن هایی که بهترین عملکرد را در مجموعه اعتبارسنجی دارند، بازیابی می شوند.

در ادامه مدل را با تابع هزینه ی `categorical_crossentropy` و `dice_loss` (که خودمان از پایه کد کردیم) با بهینه ساز آدام آموزش دادیم. معیار ارزیابی در طول آموزش را هزینه، دقت و `mean_iou` (که خودمان از پایه کد کردیم) گذاشتیم.

نهایتاً مدل در ۲۰ دوره با بیج سائز ۱۶ و ابعاد داده ی ورودی ۳۶۸\*۶۴۰ آموزش دید.

## ۴-۴) ارزیابی و تحلیل نتایج

### mIoU

`Mean Intersection over Union (mIoU)` یک معیار پرکاربرد در ارزیابی عملکرد مدل های بینایی کامپیوتر است. این شاخص دقت محلی سازی شی را با محاسبه نسبت بین تقاطع و اتحاد مناطق حقیقت پیش بینی شده و زمینی برای هر کلاس اندازه گیری می کند، سپس میانگین این مقادیر را در همه کلاس ها محاسبه می کند.

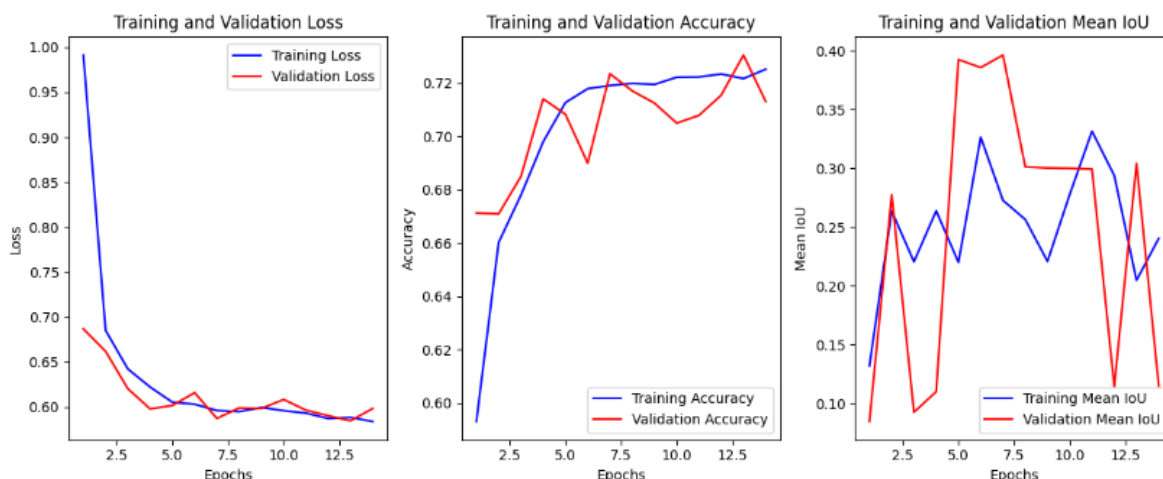
برای درک `mIoU`، می توان سناریویی را در نظر گرفت که در آن یک تصویر به چندین کلاس تقسیم می شود (به عنوان مثال، ماهی، غواص، قایق). برای هر کلاس، `IoU` با تقسیم مساحت همپوشانی بین ماسک های حقیقت پیش بینی شده و زمینی بر مساحت اتحاد آنها محاسبه می شود. `IoU` از ۰ تا ۱ متغیر است، جایی که مقدار نزدیکتر به ۱ نشان دهنده همپوشانی بهتر بین پیش بینی و حقیقت پایه است.

بعد از محاسبه مقادیر IoU برای همه کلاس‌های یک تصویر، mIoU با میانگین‌گیری این مقادیر به دست می‌آید. این متریک اندازه‌گیری تلفیقی از دقت تقسیم‌بندی در تمام کلاس‌ها را ارائه می‌دهد و بینشی را در مورد اینکه مدل چگونه اشیاء مختلف را در یک تصویر بومی سازی می‌کند، ارائه می‌دهد. mIoU بالاتر نشان دهنده عملکرد بخش‌بندی بهتر است، در حالی که مقدار پایین‌تر نشان دهنده عملکرد ضعیف‌تر یا عدم دقت در ترسیم شیء است.

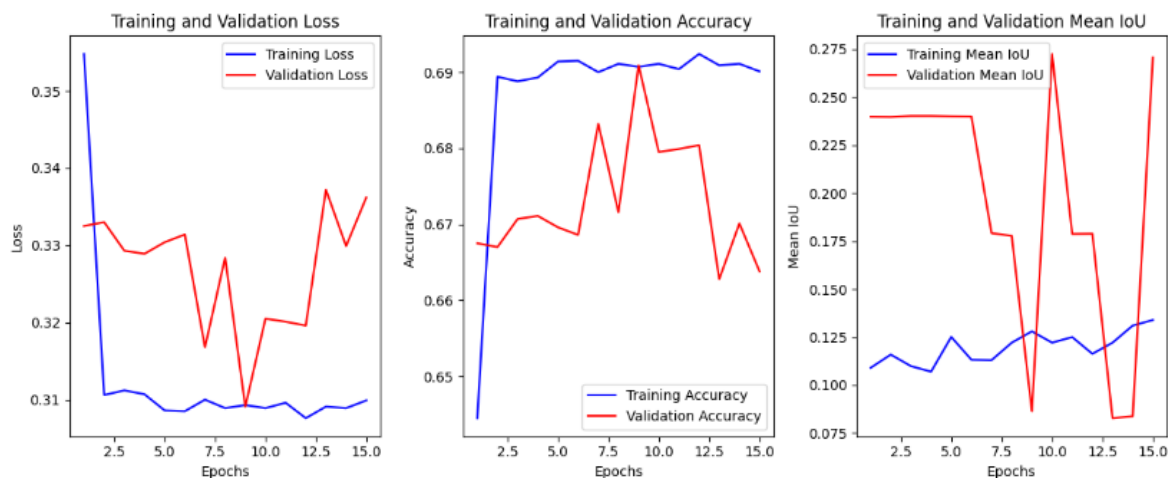
در عمل، mIoU به ویژه در ارزیابی عملکرد کلی مدل‌های تقسیم‌بندی در مقایسه با IoU فردی برای هر کلاس مفید است. این یک مقدار عددی واحد را ارائه می‌دهد که می‌تواند برای مقایسه مدل، تنظیم پارامترها و معیار استفاده شود.

در ادامه به گزارش عملکرد مدل‌هایی که در طول انجام پروژه یاد دادیم، می‌پردازیم:

مدل (۱) Unet با تابع هزینه categorical\_crossentropy بدون data augmentation



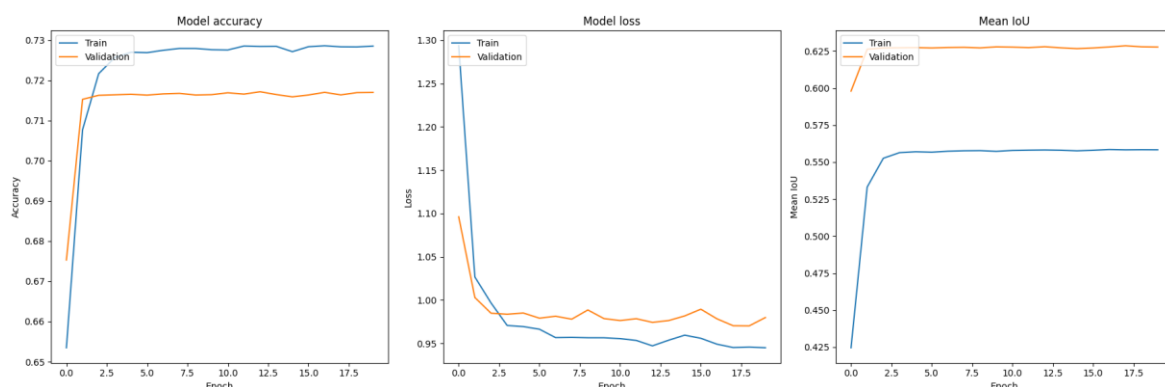
مدل (۲) Unet با تابع هزینه dice\_loss بدون data augmentation



ما برای نشان دادن تاثیر استفاده از داده افزایی و تابع هزینه ترکیبی، مدل ۱ و ۲ را اجرا کردیم. همانطور که از گزارش این دو نمودار مشخص است، قدرت طبقه بندی خوبی ندارند. با این وجود در مقایسه ی دو مدل می توان فهمید که استفاده از categorical\_crossentropy دقت بالا و هزینه پایین به همراه خواهد داشت. از طرف دیگر، با استفاده از هزینه dice\_loss، می توان به داده های نامتوازن تا حدی غلبه کرد. (همانند توضیحات ارائه شده در بخش ۱)

در ادامه به ترکیب دو تابع هزینه پرداختیم.

### مدل اصلی (۱) UNET



شکل ۱۳ UNET

منحنی loss تمرین از حدود ۱.۲ شروع می شود و در طول دوره ها به طور پیوسته کاهش می یابد و در پایان تمرین به حدود ۰.۹۵ می رسد. این نشان می دهد که مدل یاد می گیرد تا داده های آموزشی را به خوبی برازش دهد. منحنی ضرر اعتبار نیز از حدود ۱.۲ شروع می شود و در طول دوره ها کاهش می یابد، اگرچه بیشتر از منحنی تلفات تمرینی نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

منحنی دقت از حدود ۰.۶۶ شروع می شود و در طول دوره ها به طور پیوسته افزایش می یابد و در پایان تمرین به حدود ۰.۷۲ می رسد. این نشان می دهد که مدل در حال یادگیری طبقه بندی صحیح پیکسل ها در تصاویر آموزشی است. منحنی صحت اعتبارسنجی نیز از حدود ۰.۶۶ شروع می شود و در طول دوره ها افزایش می یابد، اگرچه بیشتر از منحنی دقت تمرین نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

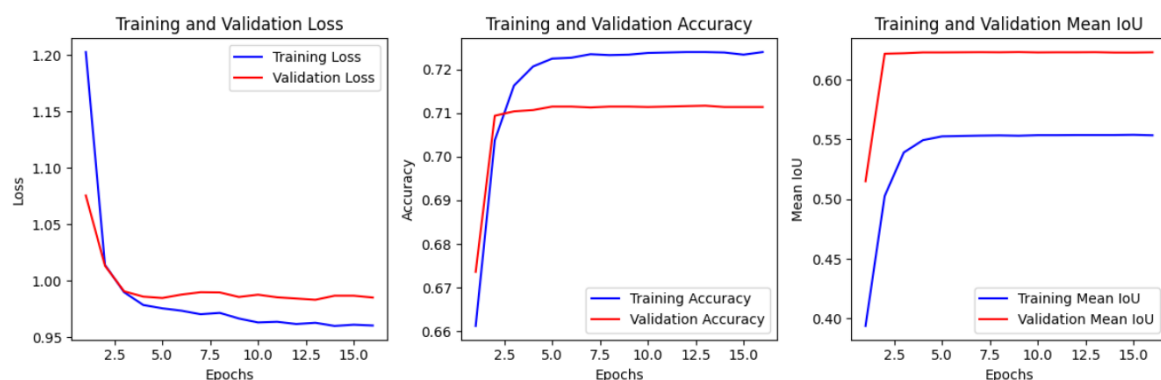
منحنی میانگین IoU تمرین از حدود ۰.۴۰ شروع می شود و به طور پیوسته در طول دوره ها افزایش می یابد و در پایان تمرین به حدود ۰.۵۵ می رسد. IoU یا تقاطع روی اتحاد، معیاری است که میزان شباهت ماسک تقسیم بندی پیش بینی شده را با ماسک حقیقت زمین اندازه گیری می کند. بنابراین، میانگین

بالا تر IoU به این معنی است که مدل در بخش بندی اشیاء در تصاویر بهتر عمل می کند. منحنی میانگین اعتبارسنجی IoU نیز از حدود ۰.۴۰ شروع می شود و در طول دوره ها افزایش می یابد، اگرچه بیشتر از منحنی میانگین آموزش IoU نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

به طور کلی، طرح نشان می دهد که مدل در حال یادگیری تقسیم بندی اشیاء در مجموعه داده SUIM به خوبی است. منحنی های کاهش و دقت تمرین به ترتیب در حال کاهش و افزایش هستند، که نشان می دهد مدل به خوبی با داده های آموزشی برازش می کند. منحنی های از دست دادن اعتبار و دقت نیز روندهای مشابهی را دنبال می کنند، که نشان می دهد مدل به خوبی به داده های دیده نشده تعمیم می یابد. میانگین منحنی های IoU نیز در حال افزایش است، که نشان می دهد مدل اشیاء موجود در تصاویر را با دقت بیشتری تقسیم می کند.

## مدل اصلی ۲) TA-UNET

گزارش آموزش مدل در ۱۷ اپاک با استفاده از early stopping را در زیر مشاهده می کنید:



شکل ۱۴ TAUNET -

منحنی loss از حدود ۱.۲ شروع می شود و در طول دوره ها به طور پیوسته کاهش می یابد و در پایان تمرین به حدود ۰.۹۵ می رسد. این نشان می دهد که مدل یاد می گیرد تا داده های آموزشی را به خوبی برازش دهد. منحنی ضرر اعتبار نیز از حدود ۱.۲ شروع می شود و در طول دوره ها کاهش می یابد، اگرچه بیشتر از منحنی تلفات تمرینی نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

منحنی دقت از حدود ۰.۶۶ شروع می شود و در طول دوره ها به طور پیوسته افزایش می یابد و در پایان تمرین به حدود ۰.۷۲ می رسد. این نشان می دهد که مدل در حال یادگیری طبقه بندی صحیح پیکسل ها در تصاویر آموزشی است. منحنی دقت اعتبارسنجی نیز از حدود ۰.۶۶ شروع می شود و در طول دوره

ها افزایش می یابد، اگرچه بیشتر از منحنی دقت آموزش نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

منحنی میانگین IoU از حدود ۰.۴۰ شروع می شود و به طور پیوسته در طول دوره ها افزایش می یابد و در پایان آموزش به حدود ۰.۵۵ می رسد. IoU یا تقاطع روی اتحاد، معیاری است که میزان شباهت ماسک تقسیم بندی پیش بینی شده را با ماسک حقیقت زمین اندازه گیری می کند. بنابراین، میانگین بالاتر IoU به این معنی است که مدل در بخش بندی اشیاء در تصاویر بهتر عمل می کند. منحنی میانگین اعتبارسنجی IoU نیز از حدود ۰.۴۰ شروع می شود و در طول دوره ها افزایش می یابد، اگرچه بیشتر از منحنی میانگین آموزشی IoU نوسان دارد. این نشان می دهد که مدل به خوبی به داده های دیده نشده تعمیم می یابد.

به طور کلی، پلات این مدل نشان می دهد که مدل در حال یادگیری خوب تقسیم بندی اشیاء در مجموعه داده SUIM است. منحنی های هزینه و دقت به ترتیب در حال کاهش و افزایش هستند، که نشان می دهد مدل به خوبی با داده های آموزشی برازش می کند. منحنی های از دست دادن اعتبار و دقت نیز روندهای مشابهی را دنبال می کنند، که نشان می دهد مدل به خوبی به داده های دیده نشده تعمیم می یابد. میانگین منحنی های IoU نیز در حال افزایش است، که نشان می دهد مدل اشیاء موجود در تصاویر را با دقت بیشتری تقسیم می کند.

هر دو  $loss$  اعتبار و دقت اعتبارسنجی کاهش/افزایش (به ترتیب) در طول دوره ها، نشان می دهد که مدل از داده های آموزشی یاد می گیرد. با این حال، منحنی  $loss$  اعتبارسنجی معمولاً بیشتر از منحنی دقت اعتبار نوسان دارد. این به این دلیل است که  $loss$  بر روی طبقه بندی پیکسل های منفرد تمرکز می کند، در حالی که دقت ممکن است برخی از خطاها را میانگین کند. نوسانات از دست دادن نشان می دهد که مدل هنوز در حال تنظیم به داده های دیده نشده در مجموعه اعتبارسنجی است.

میانگین IoU نیز در طول دوره ها افزایش می یابد، اما با نوسانات کمتر در مقایسه با از دست دادن اعتبار. این نشان می دهد که این مدل به طور کلی توانایی خود را برای تقسیم بندی دقیق اشیاء در مجموعه اعتبارسنجی بهبود می بخشد. افزایش نرم تر IoU در مقایسه با دقت می تواند نشان دهد که مدل ممکن است به درستی برخی از پیکسل ها را به صورت جداگانه طبقه بندی کند (که منجر به دقت بالاتر می شود) اما در تلاش برای ثبت دقیق اشکال کلی اشیاء است.

به طور کلی در حالی که هم دقت و هم IoU منعکس کننده یادگیری مثبت هستند، IoU تصویر ظریف تری از اینکه مدل چگونه کل اشیاء را تقسیم می کند ارائه می دهد. افزایش هموارتر در میانگین اعتبارسنجی IoU نشان می دهد که این مدل ممکن است دقت تقسیم بندی کلی شی را بر طبقه بندی پیکسل های فردی در داده های دیده نشده اولویت دهد.

مدل انتخابی نهایی ما، مدل UNET است. البته این انتخاب بسته به دیتاست موجود است. در این دیتاست SUIM مدل UNET با پیچیدگی و افزونگی کمتر توانسته تفاوت عملکرد بهتری داشته باشد. البته که این تفاوت عملکرد خیلی زیاد نیست اما بدلیل افزونگی کمتر، مدل بهتر اعلام می شود.