



فصلنامه علمی پژوهش‌های سایبرالکترونیک

صاحب امتیاز: دانشکده علوم و مهندسی دفاعی دانشگاه افسری و تربیت پاسداری امام حسین (ع)

سر دبیر: دکتر محمدعلی جوادزاده

مدیر مسئول: دکتر محمد کاظم بصیرتی

اعضای هیئت تحریریه

دکتر محمد کاظم بصیرتی	استادیار دانشگاه
دکتر حمیدرضا لشکریان	استادیار دانشگاه
دکتر مسعود باقری	دانشیار دانشگاه
دکتر محمدعلی جوادزاده	استادیار دانشگاه
دکتر حمید باباخانیان	مدرس دانشگاه
دکتر ضرغام رستمی	استادیار دانشگاه
دکتر حسین صابری	استادیار دانشگاه
دکتر مهدی علی‌نژاد	استادیار دانشگاه
مهندس سید محمود قطبانی	مدرس دانشگاه
مهندس علیرضا صفایی	مدرس و عضو هیئت علمی دانشگاه
مهندس حامد شیروی	مدرس دانشگاه
مهندس مرتضی یزدان‌پناه	مدرس دانشگاه
مهندس یوسف برجعلی	مدرس دانشگاه



همکاران علمی و داوران این شماره

دکتر منصور احدنژاد، دکتر حمید اکبری، دکتر رضا جلالی، دکتر بهروز خادم، دکتر علی کریمی، دکتر مصطفی عباسی،
دکتر مهدی علی‌نژاد، دکتر مهدی نقوی، مهندس علی بهتری، مهندس حامد شیروی، مهندس رامین دلیر،
مهندس سامان کشوری، مهندس پیمان قنوازی، مهندس مجید حاج جواد

مدیر اجرایی: مهندس علیرضا صفایی

ویراستار ادبی: سرکار خانم مهندس طوفانی

ویراستار انگلیسی و مترجم: مهندس فرامرز شمس سوسه‌باب

حروفچین و صفحه‌آرا: مهندس مرتضی یزدان‌پناه

طراح جلد: مهندس علی ذوالفقاریان

شمارگان: ۱۰۰ (نسخه چاپی) ۲۰۰ (نسخه الکترونیک) قیمت نسخه چاپی:

نشانی: تهران، بزرگراه شهید بابایی، بعد از پل لشکرک، دانشگاه افسری و تربیت پاسداری امام حسین (ع)، دانشکده علوم و مهندسی دفاعی، تلفن: ۷۷۱۰۳۰۵۰ داخلی ۱۰۹

Web Site: www.Jce.ihuo.ac.ir

E-mail: Jce@ihuo.ac.ir

تمام حقوق مادی و معنوی برای دانشکده علوم و مهندسی دفاعی محفوظ است.

فهرست

- ❖ **سخن سردیر**..... (یازده)
- ❖ **تقسیم‌بندی معنایی برای تصاویر پهپاد با استفاده از روش‌های یادگیری عمیق**..... ۱۳
 - **سید محمدعلی ضیائی**
- ❖ **کاربرد سطوح انتخاب‌گر فرکانسی پهن باند در الکترونیک دفاعی**..... ۲۷
 - **علی عزیزی**
- ❖ **بررسی ارتباطات بی‌سیم برای شبکه هوشمند مبتنی بر رویکردها و فناوری‌های نوپدید از دیدگاه EMC**..... ۳۷
 - **ضرغام رستمی، احمد عطائی**
- ❖ **تحلیل و شبیه‌سازی ساختار نیم جمع‌کننده‌های نوری مبتنی بر بلورهای فوتونی و ارائه پیشنهادهایی جهت بهبود عملکرد در مدارهای مجتمع نوری**..... ۴۷
 - **محمد مقصودی**
- ❖ **مزایای استفاده از جنگ سایبر الکترونیک در محیط‌های عملیاتی**..... ۵۷
 - **مهرداد نجفی نوکاشتی، محمدرضا موحدی‌صفت**
- ❖ **ارائه یک دسته‌بندی بر اساس مشخصه‌های سیستم عامل‌های بی‌درنگ**..... ۷۱
 - **مهدی قزلسفلو، مهدی نقوی**

تقسیم‌بندی معنایی برای تصاویر پهباد با استفاده از

روش‌های یادگیری عمیق

سید محمدعلی ضیائی^۱

چکیده

در این مقاله، به تقسیم‌بندی معنایی برای تصاویر پهباد به منظور شناسایی اجسام موجود در تصاویر پرداخته شده است. مجموعه داده‌های تقسیم‌بندی معنایی پهباد با وضوح بالا به عنوان مکمل، چالش‌های جدیدی از جمله تنوع در مقیاس بزرگ، تشخیص شی متحرک و حفظ سازگاری زمانی را به همراه دارد. مجموعه داده‌های ارائه شده بر روی افزایش ایمنی پرواز و شناسایی اجسام در مناطق شهری تمرکز دارد. چندین روش پایه یادگیری عمیق بر روی مجموعه داده ارائه شده است. در میان این روش‌ها، **Multi-Scale-Dilation** پیشنهادی از طریق استخراج ویژگی‌های **multi-scale** دارد که دارای بهترین عملکرد است. میانگین امتیاز آن به ۹۶٫۷۵ درصد می‌رسد که بیش از ۱٫۶ درصد عملکرد بهتری نسبت به سایرین دارد. همچنین تأثیر منظم‌سازی مکانی و زمانی برای دنباله‌ای از داده‌ها را با استفاده از بهینه‌سازی فضای ویژگی بررسی کرده‌ایم.

کلیدواژه‌ها: تقسیم‌بندی معنایی تصاویر، مجموعه داده‌های پهباد، شناسه پهباد، تصاویر پهباد، یادگیری عمیق

۱- مقدمه

درک صحنه بصری که به عنوان یک پایه درک، برای بسیاری از زمینه‌ها مانند رباتیک و رانندگی خودکار عمل می‌کند، در سال‌های اخیر پیشرفت بسیار زیادی کرده است. مؤثرترین و موفق‌ترین روش‌ها برای درک صحنه، الگوریتم‌های یادگیری عمیق را به عنوان سنگ بنای خود می‌پذیرند، زیرا می‌تواند دانش معنایی سطح بالایی را از داده‌های آموزشی استخراج کند. با این حال، چالش این است که یادگیری عمیق به تعداد زیادی نمونه برای آموزش نیاز دارد تا دانش مفید را به جای نویز یاد بگیرد. برای برنامه‌های کاربردی در دنیای واقعی، تقسیم‌بندی معنایی به عنوان بخشی از درک صحنه، به این صورت است که برای هر پیکسل در تصویر، برچسبی را اختصاص می‌دهد. برای استفاده بهتر از روش‌های یادگیری عمیق، تعداد زیادی عکس با برچسب مورد نیاز است. در حال حاضر، چندین مجموعه داده تقسیم‌بندی معنایی عمومی وجود دارد که فقط بر روی اشیاء رایج در تصاویر طبیعی تمرکز می‌کنند. همه آن‌ها از روی زمین عکس می‌گیرند که تصاویر با دوربین‌های نصب شده روی وسایل نقلیه گرفته می‌شود. اشیاء مورد توجه شامل عابران پیاده، اتومبیل‌ها، جاده‌ها، خطوط، چراغ‌های راهنمایی، درختان و سایر اشیاء اطراف در نزدیکی خیابان‌ها هستند. با توجه به پلتفرم‌های سنجش از راه دور، تعداد مجموعه‌های داده برای تقسیم‌بندی معنایی بسیار کم است و تصاویر اغلب در نمای نادری گرفته می‌شوند که در آن فقط بالای اشیاء قابل مشاهده است. برای تصاویر هوایی همه مجموعه داده‌های هدف، بر توسعه روش‌های جدید تقسیم‌بندی معنایی فعلی تأکید داشته‌اند.

علی‌رغم کاربرد وسیع هواپیمای بدون سرنشین (UAV) برای رصد زمین، متأسفانه تحقیقات کمی برای تجزیه و تحلیل صحنه شهری انجام شده است. تحقیق تقسیم‌بندی معنایی برای صحنه‌های شهری می‌تواند پایه و اساس برنامه‌هایی مانند نظارت بر ترافیک، تصادفات رانندگی، نظارت بر تراکم جمعیت و نظارت بر فضای سبز شهری باشد. اگرچه مجموعه داده‌هایی برای تشخیص، ردیابی و تحلیل رفتار وجود دارد، اما در این مقاله، مجموعه داده‌های تقسیم‌بندی معنایی پهباد جدید با تصاویر پهباد با وضوح بالا در نماهای مایل ارائه شده است که برای تقسیم‌بندی معنایی صحنه‌های شهری طراحی شده است. چندین چالش را برای مجموعه داده جدید مطرح کرده‌ایم: تنوع در مقیاس بزرگ بین اشیاء در فواصل مختلف یا دسته‌های مختلف، تشخیص جسم متحرک در صحنه خیابان شهری، و حفظ ثبات زمانی برای پیش‌بینی بهتر در فریم‌ها. این چالش‌ها منحصر به فرد بودن مجموعه داده ما را نشان خواهد داد که در مجموع، ۴۰۰ تصویر از نمای چشم پرنده (دوربین) است که شامل بیش از ۲۰ خانه که در ارتفاع ۵ تا ۳۰ متری از سطح زمین است به دست آمده است. اندازه مجموعه داده ما ده برابر مجموعه داده [۲] Vaihingen، پنج برابر مجموعه داده [۳] CamVid و دو برابر مجموعه داده [۲] Potsdam در مورد تعداد پیکسل‌های برچسب‌گذاری شده است. برچسب ارائه شده، با FPS 20 به دست می‌آیند، و تصاویر بدون برچسب نیز با FPS 20 برای کاربران ارائه می‌شود. تصاویر اضافی برای کمک به تشخیص بالقوه شی ارائه شده است. برای کار برچسب‌گذاری معنایی و آزمایش قابلیت استفاده از مجموعه داده، از چندین شبکه عصبی عمیق معمولی (DNN) استفاده می‌شود، از جمله [۴] FCN-8s، شبکه [۵] Dilation و [۶] U-Net که به طور گسترده برای کار تقسیم‌بندی

معنایی در مجموعه داده‌های مختلف استفاده شده که پایدار هستند. در ادامه، این مقاله به شرح زیر سازمان‌دهی شده است: بخش ۲ به کارهای مرتبط پرداخته شده است بخش ۳ نحوه ساخت مجموعه داده‌های پهپاد را برای تقسیم‌بندی معنایی صحنه شهری، از جمله مشخصات داده‌ها، تعریف کلاس، روش‌های حاشیه‌نویسی و تقسیم مجموعه داده‌ها توضیح می‌دهد. بخش ۴ وظیفه برچسب‌گذاری معنایی را برای مجموعه داده شناسه پهپاد ارائه می‌کند. این بخش شامل تصویر کار و روش‌های پایه برای کار است. بخش ۵ نتایج آزمایش مربوطه را با تجزیه و تحلیل برای روش‌های پایه نشان می‌دهد. بخش ۶ مقایسه‌ای نسبت به کارهای انجام شده صورت می‌گیرد در نهایت، بخش ۷ نکات پایانی و چشم‌انداز مجموعه داده شناسه پهپاد را ارائه می‌دهد.

۲- پیشینه تحقیق

تقسیم‌بندی معنایی، تصاویر را به مناطقی تقسیم کرده و هر منطقه را با یک برچسب کلاس از پیش تعریف شده اختصاص می‌دهد و سپس برچسب‌گذاری می‌کند. شناسایی طرح صحنه، اطلاعاتی در مورد توزیع فضایی شی و ارتباط آن‌ها با محیط ارائه می‌دهد. مروری کوتاه بر تقسیم‌بندی معنایی در زیر ارائه شده است.

نام نویسنده	عنوان مقاله	سال انتشار	نتایج
Girisha S (۱) Ujjwal Verma (۲) Pai M. M (۳) Radhika M. Pai (۴)	تجزیه و تحلیل عملکرد الگوریتم‌های تقسیم‌بندی معنایی برای مجموعه داده‌های ویدئویی هوایی پهپاد جدید با حاشیه‌نویسی دقیق (ManipalUAVid)	۲۰۱۹	در این مقاله از مجموعه داده‌های ویدئویی هوایی ایجاد شده به عنوان بخشی از این کار برای تجزیه و تحلیل عملکرد چهار رویکرد استاندارد برای تقسیم‌بندی معنایی یعنی میدان تصادفی شرطی (CRF)، معماری FCN-8 و U-Net و DeepLabV3C استفاده شده است. معماری برای رویکرد مبتنی بر CRF ویژگی طرح بافت همراه با رنگ، مکان و پتانسیل لبه گنجانده شده است. این رویکردها برای تقسیم‌بندی معنایی با دقت پیکسل (CRF، ۰.۹۳)، (U-Net) (FCN، ۰.۹۲) و (DeepLabV3C، ۰.۹۱) دست یافته‌اند که به صورت رقابتی عمل می‌کنند.
Shuanglong Liu (۱) Wayne Luk (۲)	به‌سوی یک شتاب‌دهنده کارآمد برای تقسیم‌بندی تصویر سنجش از راه دور مبتنی بر DNN در FPGA	۲۰۱۹	این مقاله یک سیستم شتاب‌دهنده را برای پردازش لحظه‌ای روی برد برای تقسیم‌بندی RSI در پلتفرم‌های فضایی پیشنهاد می‌کند. یک معماری یکنواخت را پیشنهاد می‌کند که می‌تواند به‌طور مؤثر لایه‌های Conv و Deconv را در مدل‌های DNN پیاده‌سازی کند. این معماری با بهره‌برداری از سطوح مختلف موازی‌سازی، روش همجوشی لایه‌ها و استفاده کامل از DSP‌ها بهینه‌سازی می‌شود.
Fahad LATEEF (۱) Yassine RUICHEK (۲)	بررسی تقسیم‌بندی معنایی با استفاده از تکنیک‌های یادگیری عمیق	۲۰۱۹	در این مقاله، یک بررسی جامع از تکنیک‌های یادگیری عمیق مورد استفاده برای تقسیم‌بندی معنایی ارائه شده است. روش‌های بررسی شده با توجه به مفهوم رایج زیربنای معماری آن‌ها در ده کلاس طبقه‌بندی شده‌اند. در این مقاله همچنین خلاصه‌ای از این روش‌ها ارائه شده است که برای هر یک از آن‌ها، ایده اصلی، منشأ معماری، معیارهای آزمایش، دردسترس بودن کد و سال انتشار را بیان کرده‌اند. سی‌وپنج مجموعه داده که این روش‌ها بر روی آن‌ها اعمال شده است گزارش شده و با جزئیات شرح داده شده است که ماهیت محیط، تعداد کلاس‌ها، وضوح، تعداد تصاویر و روشی را که بهترین عملکرد را در هر کدام از آن‌ها تا به امروز به بهترین نحو کسب کرده است.
Shervin Minaee (۱) Yuri Boykov (۲) Fatih Porikli (۳) Antonio Plaza (۴)	مروری تقسیم‌بندی تصویر با استفاده از یادگیری عمیق	۲۰۲۱	در این مقاله الگوریتم‌های تقسیم‌بندی تصویر را بر اساس مدل‌های یادگیری عمیق مورد بررسی قرار گرفت است که عملکرد چشمگیری در وظایف و معیارهای تقسیم‌بندی تصویر مختلف داشته‌اند. مدل‌های مبتنی بر یادگیری عمیق از جمله

CNN و CNN dilated, R-CNN, RNN, FCN بر روی مجموعه داده‌های PASCAL VOC, MS COCO, Cityscapes, ADE و k₂ بررسی کرده‌اند.			۵) Nasser Kehtarnavaz ۶) Demetri Terzopoulos
در این مقاله یک معماری CNN مبتنی بر رمزگذار- رمزگشا برای تقسیم معنایی ویدئوهای هوایی پهپاد ارائه می‌کند. معماری پیشنهادی از یک رمزگذار جدید متشکل از دوشاخه رمزگذاری موازی با دو فریم کلیدی متوالی ویدئو به عنوان ورودی شبکه استفاده می‌کند. با ادغام ویژگی‌های استخراج شده از دوشاخه رمزگذاری، شبکه می‌تواند اطلاعات زمانی را بیاموزد که نیاز به یک مازول متوالی اضافی را از بین می‌برد. علاوه بر این، از یک مازول پالایش کننده ویژگی در مسیر رمزگشا استفاده می‌کند. این مازول مرزهای تقسیم بندی صاف تری ایجاد می‌کند. معماری پیشنهادی به دقت ۰.۷۹ در مجموعه داده ManipalUAVid دست یافت.	۲۰۲۱	تقسیم بندی معنایی پیشرفته و ویدئوهای هوایی پهپاد با جاسازی اطلاعات زمانی	۱) Girisha S ۲) Ujjwal Verma ۳) Pai M. M ۴) Radhika M. Pai

جدول ۱: پیشینه تحقیق

۳- مجموعه داده

طراحی یک مجموعه داده پهپاد مستلزم تفکر دقیق در مورد استراتژی جمع آوری داده، پروتکل پرواز پهپاد و انتخاب کلاس شی برای حاشیه نویسی است. کل فرایند با در نظر گرفتن سودمندی و اثربخشی برای تحقیق تقسیم بندی معنایی پهپاد طراحی شده است که ما از آن استفاده می‌کنیم. در این بخش، نحوه ایجاد مجموعه داده از تصاویر گرفته شده توسط پهپاد نشان داده شده است. بخش ۱-۳ راهبرد کسب داده را نشان می‌دهد. بخش ۳-۳ و ۳-۴ به ترتیب کلاس های کار و روش های حاشیه نویسی را شرح می‌دهد. بخش ۵-۳ تقسیم بندی داده ها را برای وظیفه تقسیم بندی معنایی نشان می‌دهد.

مشخصات داده‌ها

روش کسب داده و حاشیه نویسی ای برای تقسیم بندی معنایی پهپاد در صحنه های شهری پیچیده طراحی شده است که دارای قابلیت تشخیص شی ثابت و متحرک است. ویژگی های زیر برای مجموعه داده ها در نظر گرفته شده است که به شرح زیر است.

نمای مایل: برای پلتفرم پهپاد، ضبط تصاویر یا ویدئوها در یک نمای مایل یا در نمای نادر طبیعی است. نماهای نادر تصاویر ماهواره ای رایج است؛ زیرا فاصله بین دوربین تا زمین زیاد است. نماهای نادر برای نمایش اشیاء، در تصویر تغییرناپذیری ایجاد می‌کند، زیرا فقط قسمت بالای اشیاء را می‌توان مشاهده کرد. با این حال، نمایش محدود همچنین باعث سردرگمی در تشخیص شی در بین اشیاء مختلف می‌شود. در مقابل، یک نمای مایل نمایش متنوعی از اشیاء با زمینه صحنه غنی را ارائه می‌دهد که می‌تواند برای تشخیص شی مفید باشد. هنگامی که پهپاد نزدیک تر به زمین پرواز می‌کند، می‌توان ناحیه بزرگ تری با جزئیات بیشتر مشاهده کرد که باعث ایجاد تنوع در مقیاس بزرگ در یک تصویر می‌شود. برای مشاهده در نمای مایل، زاویه دوربین در حدود ۶۵ درجه نسبت به جهت عمودی تنظیم می‌شود.

وضوح بالا: حالت ضبط ویدئو با وضوح K4 را با ارتفاع پرواز ایمن در حدود ۵۰ متر اتخاذ شده است. توجه داشته باشید قالب تصاویر ضبط شده در ارتفاع ۳۰ متری از سطح زمین است. وضوح تصویر ۲۱۶۰×۴۰۹۶ یا ۴۰۰۰×۶۰۰۰ است. در این تنظیمات که از نظر بصری تصویر به اندازه کافی واضح است برای متمایز کردن بیشتر اشیاء، اجسامی که به صورت افقی دور هستند نیز قابل شناسایی هستند. علاوه بر این، حتی می‌توان انسان‌هایی را که در نزدیکی پهپاد قرار دارند، شناسایی کرد.

برچسب زدن متوالی: مجموعه داده موجود ما برای وظیفه تقسیم‌بندی معنایی طراحی شده است که ترجیح می‌دهیم تصاویر را در یک دنباله برچسب‌گذاری کنیم؛ جایی که بتوان ثبات پیش‌بینی را ارزیابی کرد. از آنجایی که برچسب‌گذاری متراکم در فضای زمانی بسیار پرهزینه است روش ارائه شده توسط ما به این صورت است که ۱۰ تصویر را با فاصله ۵ ثانیه در هر دنباله برچسب‌گذاری می‌کنیم.

صحنه‌های پیچیده و پویا با اشیاء متنوع: مجموعه داده مورد استفاده ما با هدف دستیابی به پیچیدگی صحنه در دنیای واقعی طراحی شده است. به طور مثال صحنه‌های نزدیک خیابان‌ها برای مجموعه داده‌های شناسه پهپاد انتخاب شده‌اند؛ زیرا به اندازه کافی پیچیده هستند و فعالیت‌های انسانی پویاتری دارند. اشیاء مختلفی در صحنه ظاهر می‌شوند؛ مانند اتومبیل‌ها، عابران پیاده، ساختمان‌ها، جاده‌ها، پوشش گیاهی و غیره.

تنوع داده‌ها: در مجموع، ۳۰ توالی ویدئویی کوچک پهپاد در ۲۰ مکان مختلف گرفته شده است. جمع‌آوری داده‌ها در شرایط آب‌وهوایی خوب با روشنایی کافی انجام شد. داده‌های به دست آمده در محیط‌های تاریک یا سایر شرایط آب‌وهوایی مانند بارش برف یا باران به تکنیک‌های پردازش خاصی نیاز دارند که مجموعه داده‌های فعلی ما بر این مورد تمرکز ندارد.

DJI phantom3 pro و DJI phantom4 پهپادهای سبک‌وزن مدرنی هستند که برای جمع‌آوری داده‌ها مورد نیاز ما استفاده می‌شود است. پهپادها به طور پیوسته با حداکثر سرعت پرواز ۱۰ متر بر ثانیه پرواز می‌کنند. دوربین‌های پیش‌فرض نصب شده بر روی پهپادها برای دریافت ویدئو فقط با کانال‌های RGB استفاده می‌شوند.

اندازه مجموعه داده

شناسه مجموعه داده‌های پهپاد ما دارای ۴۰۰ تصویر و هر کدام به اندازه ۲۱۶۰×۴۰۹۶ یا ۴۰۰۰×۶۰۰۰ است. برای مقایسه اندازه مجموعه داده‌های مختلف برای تقسیم‌بندی معنایی، نه تنها تعداد تصاویر، بلکه اندازه هر تصویر را نیز باید در نظر بگیریم. یک معیار اصولی‌تر، مقایسه تعداد پیکسل‌های برچسب‌گذاری شده در کل داده‌ها است که ما چندین مجموعه داده تقسیم‌بندی معنایی شناخته شده را برای مقایسه انتخاب می‌کنیم. مجموعه داد [۳] CamVid دارای ۷۰۱ تصویر با اندازه ۷۲۰×۹۶۰ است که تنها یک پنجم مجموعه داده‌های ما از نظر تعداد پیکسل‌های برچسب‌گذاری شده است. مجموعه داده غول‌پیکر Cityscapes [۱] دارای ۵۰۰۰ تصویر با اندازه ۲۰۴۸×۱۰۲۴ است که ۴ برابر بزرگ‌تر از مجموعه داده‌های پهپاد ما هست. با این حال، بسیاری از اشیاء در تصاویر ما کوچک‌تر از آن‌ها هستند، واریانس بیشتر، شیء در همان تعداد پیکسل را در تشخیص شیء در یک درجه بیشتر جبران می‌کند. در

بخش ۶ نشان داده است که روش‌های یادگیری عمیق می‌توانند به نتایج کمی و کیفی رضایت‌بخشی برای اهداف تجربی دست یابند که قابلیت استفاده از شناسه مجموعه داده‌های پهپاد ما را بیشتر اثبات می‌کند.

تعریف کلاس و تجزیه و تحلیل آماری

برچسب زدن کامل انواع اشیاء در صحنه خیابان در یک تصویر پهپاد k4 بسیار پرهزینه است. در نتیجه فقط، رایج‌ترین و معروف‌ترین نوع اشیاء برای مجموعه داده فعلی ما برچسب‌گذاری می‌شوند. در مجموع، ۸ کلاس کلی برای تقسیم‌بندی معنایی انتخاب شده است؛ یعنی ساختمان (دیوار و پنجره و...)، جاده، درخت، پوشش گیاهی کوتاه، ماشین ساکن، ماشین متحرک، انسان، نمونه‌هایی از کلاس‌های مختلف است.

روش حاشیه‌نویسی

برای تصاویر پهپاد با وضوح بالا حاشیه‌نویسی‌های دقیق با برچسب‌گذاری متراکم ارائه داده‌ایم که همه برچسب‌ها، با ابزار برچسب‌زنی ما به دست می‌آیند. تقریباً ۱۲۰ دقیقه طول می‌کشد تا همه پیکسل‌ها در یک تصویر برچسب‌گذاری شوند. روش‌های حاشیه‌نویسی سطح پیکسل، سطح سوپر پیکسل و سطح چند ضلعی برای حاشیه‌نویس‌ها، همان‌طور که در شکل ۱ نشان داده شده است، ارائه شده است. برای حاشیه‌نویسی سطح سوپر پیکسل، در روش ما از یک استراتژی مشابه به عنوان مجموعه داده COCO-Stuff [۷] استفاده می‌شود. ابتدا از روش SLIC [۸] برای تقسیم کردن تصویر به سوپر پیکسل‌ها استفاده می‌کنیم که هرکدام گروهی از پیکسل‌ها هستند که از نظر فضایی به هم متصل هستند و ویژگی‌های مشابه مانند رنگ و بافت دارند. از آنجایی که ممکن است اشیاء همپوشانی داشته باشند، پیکسل‌های همپوشانی را به عنوان کلاسی که به دورین نزدیک‌تر است برچسب‌گذاری می‌کنیم.



شکل ۱: روش‌های حاشیه‌نویسی سمت چپ حاشیه‌نویسی سطح پیکسل، وسط حاشیه‌نویسی سطح فوق پیکسل، و سمت راست حاشیه‌نویسی سطح چندضلعی را نشان می‌دهد.

تقسیم مجموعه داده‌ها

کل ۳۰ دنباله ویدئویی با برچسب متراکم به بخش‌های آموزشی، اعتبارسنجی، و تقسیم‌بندی آزمایشی تقسیم می‌شوند. ما داده‌ها را کاملاً تصادفی تقسیم نمی‌کنیم، بلکه به گونه‌ای است که هر تقسیم‌بندی به اندازه کافی برای تنوع صحنه‌های مختلف، نماینده است. هر سه تقسیم باید شامل همه کلاس‌ها باشد. داده‌های ما در سطح سکانس تقسیم می‌شوند و هر سکانس از محل صحنه متفاوتی می‌آید. با پیروی از این طرح، ما به ترتیب ۱۵ دنباله آموزشی

(۱۵۰) تصویر برچسب‌گذاری شده) و ۵ دنباله اعتبار سنجی (۵۰ تصویر برچسب‌گذاری شده) برای تقسیم‌های آموزشی و اعتبار سنجی دریافت می‌کنیم. تقسیم آزمایشی شامل ۲۰ دنباله سمت چپ (۲۰۰ تصویر برچسب‌دار) است. نسبت اندازه بین آموزش، اعتبارسنجی و تقسیم آزمون ۳:۱:۴ است.

۴- وظیفه برچسب‌گذاری معنایی

در این بخش، وظیفه برچسب‌گذاری معنایی برای مجموعه داده ما تشریح می‌شود. بخش ۲-۴ شبکه‌های عصبی عمیق را در روش‌های پایه معرفی می‌کنند بخش ۳-۴ منظم‌سازی مکانی و زمانی را معرفی می‌کنند که عملکرد را برای همه روش‌های پایه افزایش می‌دهد.

وظیفه و متریک

وظیفه تعریف شده در شناسه مجموعه داده‌های پهپاد، پیش‌بینی برچسب‌گذاری معنایی سطح پیکسل برای تصاویر پهپاد است. عملکرد برچسب‌گذاری معنایی بر اساس معیار استاندارد IOU [۹] ارزیابی می‌شود. هدف این کار دستیابی به بالاترین میانگین امتیاز IOU تا حد امکان است. برای شناسه مجموعه داده‌های پهپاد، کلاس Clutter تعداد پیکسل نسبتاً زیادی دارد و از اشیاء معنادار تشکیل شده است که به‌جای نادیده گرفته شدن، به‌عنوان یک کلاس برای آموزش و ارزیابی در نظر گرفته می‌شود.

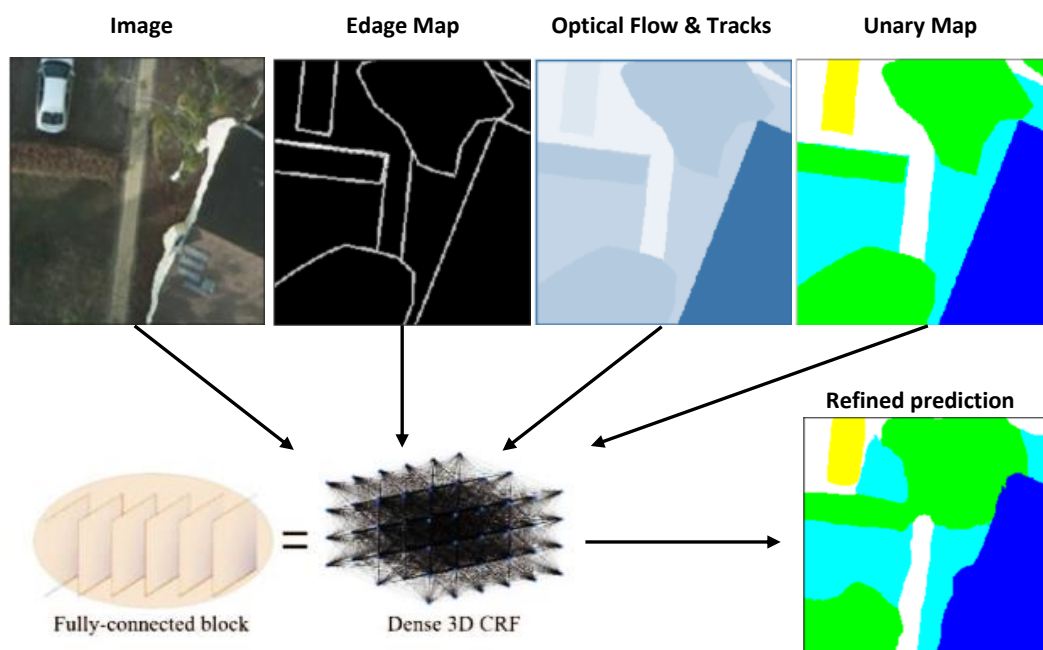
شبکه‌های عصبی عمیق برای خطوط پایه

به‌منظور ارائه مراجع عملکرد و آزمایش قابلیت استفاده از شناسه مجموعه داده‌های پهپاد ما برای کار برچسب‌گذاری معنایی، چندین مدل یادگیری عمیق را برای پیش‌بینی تصویر واحد آزمایش کرده‌ایم. اگرچه ماشین‌های ساکن و ماشین‌های متحرک را نمی‌توان از نظر ظاهری تنها از یک تصویر متمایز کرد، اما هنوز هم می‌توان بر اساس زمینه آن‌ها استنباط کرد. ما با ۳ شبکه عصبی کاملاً پیچیده عمیق و پرکاربرد شروع می‌کنیم، آن‌ها FCN-8s [۴]، شبکه Dilation [۵] و U-Net [۶] هستند. FCN-8s اغلب یک نامزد پایه خوب برای تقسیم‌بندی معنایی بوده است. این یک مدل عظیم با قابلیت استخراج ویژگی‌های قوی و مؤثر است، درعین حال ساختار ساده‌ای دارد. U-Net یک شبکه رمزگذار-رمزگشا متقارن معمولی است که در اصل برای تقسیم‌بندی در تصاویر پزشکی طراحی شده است. U-Net نیز از نظر ساختار ساده است؛ اما در حفظ مرزهای اشیاء خوب است. Multi-Scale-Dilation برای یک تصویر با وضوح بالا است که توسط یک پهپاد در یک نمای مایل گرفته می‌شود که به‌طور چشمگیری اندازه اشیاء در فواصل مختلف می‌تواند متفاوت باشد. در مجموع سه جریان برای سه مقیاس ایجاد می‌شود که برای هر جریان، یک FCN-8s اصلاح‌شده به‌عنوان ساختار اصلی استفاده می‌شود که در آن عمق برای هر بلوک کانولوشن به دلیل محدودیت حافظه کاهش می‌یابد. در اینجا، عمق فیلتر، قربانی مقیاس‌های بیشتر می‌شود. به‌منظور کاهش ازدست‌دادن جزئیات در استخراج ویژگی، لایه ادغام در بلوک کانولوشن پنجم حذف می‌شود تا یک میدان پذیرنده کوچک‌تر حفظ شود. در عوض، ویژگی‌هایی با میدان‌های دریافتی بزرگ‌تر از جریان‌های دیگر از طریق اتصال

پرش در لایه‌های conv7 به ویژگی‌های و ضوح بالاتر الحاق می‌شوند. به این ترتیب، هر جریانی استخراج ویژگی را در مقیاس خاص خود انجام می‌دهد و ویژگی‌های مقیاس‌های بزرگ‌تر برای تقویت پیش‌بینی برای جریان‌های با وضوح بالاتر تجمیع می‌شوند. مقیاس‌های چندگانه نیز ممکن است با نمونه‌برداری مستقیم تصاویر به دست آیند. با این حال، ۳ مزیت برای پردازش چند مقیاسی ما وجود دارد. ابتدا، هر پیکسل به یک دسته در فضای به عملیات دسته‌ای اختصاص داده می‌شود و تمام پیکسل‌های برچسب‌گذاری شده باید برای هر مقیاس بدون ضایعات استفاده شوند. دوم، تراز دقیقی بین جفت‌های برچسب تصویر در هر مقیاس وجود دارد، زیرا هیچ ترکیبی از پیکسل‌های تصویر و یا ترکیبی از پیکسل‌های برچسب وجود ندارد و سوم اینکه، ویژگی‌های الحاقی در لایه Conv7 نیز کاملاً در یک راستا قرار دارند. برای هر مقیاس، برچسب‌های حقیقت زمینی متناظر نیز می‌توانند از طریق عملیات فضایی به دسته‌ای به همان شیوه تولید تصاویر ورودی در جریان‌های مختلف تولید شوند. با برچسب‌های حقیقت زمینی برای هر مقیاس، آموزش عمیقاً نظارت شده می‌تواند انجام شود. تلفات در هر سه مقیاس باعث دست‌دادن آنتروپی متقاطع می‌شود.

منظم‌سازی مکانی - زمانی

برای برچسب‌گذاری معنایی، ما بیشتر این موضوع را بررسی می‌کنیم که چگونه یک نظم مکانی - زمانی می‌تواند پیش‌بینی را بهبود بخشد. استفاده از اطلاعات زمانی برای پیش‌بینی برچسب برای داده‌های توالی ارزشمند است. به طور معمول، شبکه‌های عصبی عمیق آموزش داده‌شده بر روی تصاویر منفرد نمی‌توانند پیش‌بینی‌های کاملاً منسجمی را در چندین فریم ارائه دهند. با این حال، فریم‌های مختلف مشاهداتی را از موقعیت‌های مختلف مشاهده ارائه می‌دهند که از طریق آن‌ها می‌توان سرنخ‌های متعددی برای پیش‌بینی اشیاء جمع‌آوری کرد. برای استفاده از اطلاعات زمانی در شناسه مجموعه داده‌های پهباد، از روش بهینه‌سازی فضای ویژگی (FSO) [۱۱] برای پیش‌بینی داده‌های توالی استفاده می‌کنیم. این پیش‌بینی برچسب نهایی را برای کل دنباله با اعمال CRF سه‌بعدی که هر دو حوزه مکانی و زمانی را پوشش می‌دهد، صاف می‌کند. این روش از جریان نوری و ردیابی برای پیوند دادن پیکسل‌ها در حوزه زمانی بهره می‌برد. کل پس‌پردازش به ورودی‌های داده‌های متعددی نیاز دارد، از جمله تصویر، نقشه یکپارچه از شبکه‌های عصبی عمیق، نقشه لبه، جریان نوری، و مسیرهایی که در شکل ۲ نشان داده شده است.



شکل ۲: ورودی‌های داده برای روش پس‌پردازش FSO [۱۱]. تصویر، نقشه لبه، نقشه یکپارچه، جریان نوری و مسیرها برای روش مورد نیاز هستند. نقشه لبه احتمال یال بودن هر پیکسل را نشان می‌دهد. نقاط آبی در تصویر (دومی از سمت راست) جریان نوری و مسیرها، نقاط ردیابی شده را مشخص می‌کنند. نقشه واحد، احتمالات کلاس برای هر پیکسل است که توسط شبکه‌های عصبی عمیق پیش‌بینی می‌شود.

۵- آزمایش

آزمایش‌ها ما به ۳ قسمت تقسیم شده است. ابتدا، ما نتایج تقسیم‌بندی معنایی را با آموزش شبکه‌های عصبی عمیق از ابتدا مقایسه می‌کنیم. در مرحله دوم، ما تجزیه و تحلیل می‌کنیم که چگونه مدل‌های از پیش آموزش دیده می‌توانند برای کار برجسته زدن معنایی شناسه پهپاد مفید باشند. در نهایت، ما تأثیر تنظیم مکانی - زمانی را با استفاده از روش FSO برای بخش‌بندی ویدئوی معنایی بررسی می‌کنیم. اندازه تصاویر پهپاد ما بسیار بزرگ است، در نتیجه، هر تصویر پهپاد را به ۹ تصویر کوچک‌تر با توزیع یکنواخت برش می‌دهیم که کل تصویر را برای آموزش پوشش می‌دهند. هر تصویر برش خورده در اندازه 1024×2048 است.

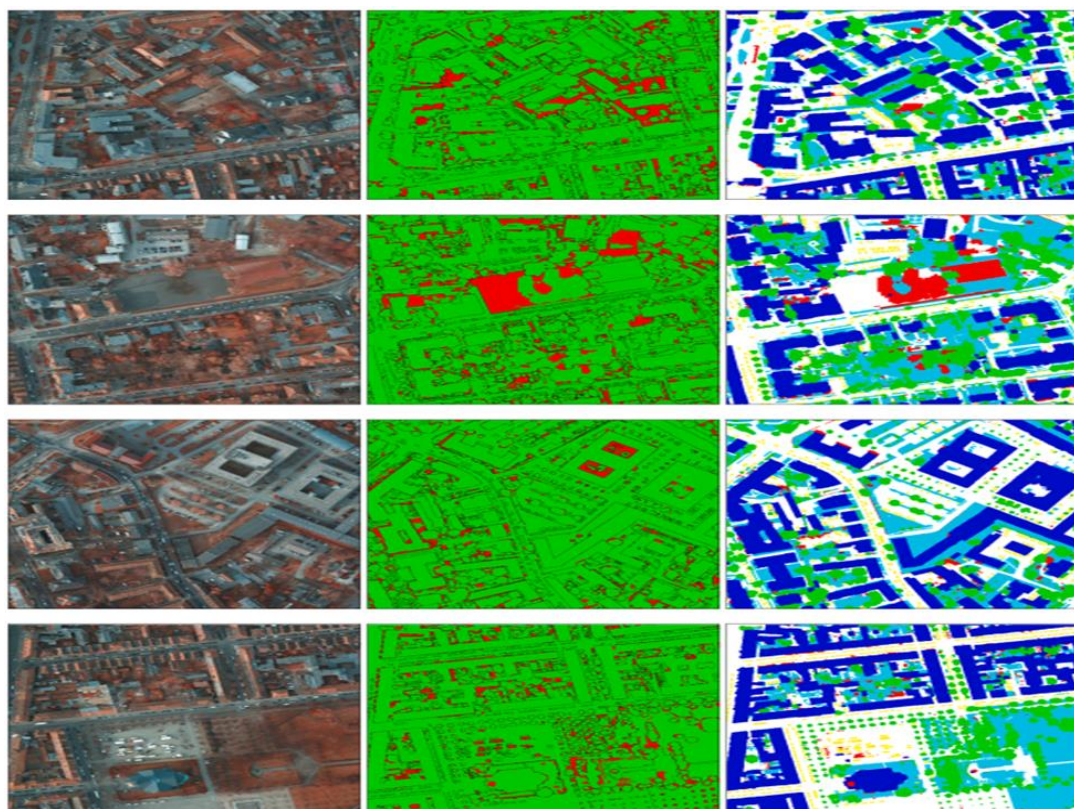
از ابتدا آموزش می‌دهیم

برای مقایسه منصفانه بین شبکه‌های مختلف، همه شبکه‌ها را با Tensorflow [۱۰] دوباره پیاده‌سازی می‌کنیم و همه شبکه‌ها با یک GPU Nvidia Titan X آموزش داده می‌شوند. به منظور قراردادن شبکه‌ها در حافظه GPU 12G، عمق برخی از لایه‌ها در شبکه U-Net، Dilation و MSDilation کاهش می‌یابد تا حداکثر در حافظه جا شود. همه مدل‌ها با بهینه‌ساز Adam برای ۲۷K تکرار (۲۰ دوره) آموزش دیده‌اند. نرخ یادگیری پایه روی ۴-۱۰ تنظیم شده است که به صورت تصاعدی به ۷-۱۰ کاهش می‌یابد. کاهش وزن برای همه وزن‌ها در هسته‌های

کانولوشن روی ۵-۱۰ تنظیم شده است. همه شبکه‌ها به صورت سرتاسر آموزش داده می‌شوند و میانگین نمرات IOU به دست آمده به درصد گزارش می‌شود. برای هر چهار شبکه، آن‌ها در تمایز طبقات ساختمان، جاده و درختان بهتر عمل می‌کنند و به امتیازات IOU بالاتر از ۵۰ درصد می‌رسند.

تنظیم دقیق مدل‌های از پیش آموزش‌دیده

برای تنظیم دقیق، همه شبکه‌ها با مجموعه داده‌های منظر شهری [۱] از قبل آموزش داده شده‌اند. داده‌های حاشیه‌نویسی دقیق از تقسیم‌بندی آموزشی و اعتبارسنجی استفاده می‌شود که در مجموع شامل ۳۴۵۰ تصویر با برچسب متراکم است. در مرحله بعد، تمام شبکه‌ها با داده‌های شناسه مجموعه داده‌های پهباد به‌خوبی تنظیم می‌شوند. نرخ یادگیری روی ۵-۱۰ تنظیم شده است که به صورت تصاعدی به ۷-۱۰ برای FCN-8 کاهش می‌یابد، و ۱۰-۴ به طور نمایی به ۷-۱۰ برای ۳ شبکه دیگر کاهش می‌یابد، زیرا آن‌ها به راحتی در حداقل محلی با نرخ یادگیری اولیه گیر می‌کنند. ۷-۱۰ در طول تمرین. جداکردن فیلترها با مقیاس‌های مختلف زمانی می‌تواند مفید باشد که اشیاء در مقیاس بزرگ ظاهر شوند. به منظور مشاهده اثر پردازش در مقیاس چندگانه، مقایسه عملکرد کیفی بین FCN-8، شبکه Dilation، U-Net و MS-Dilation Net در شکل ۳ ارائه شده است



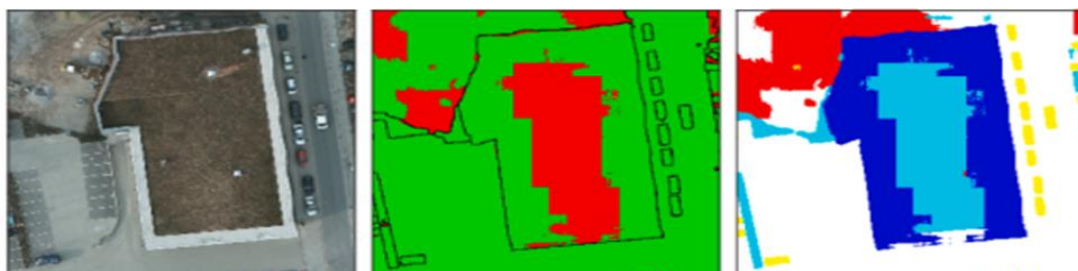
شکل ۳: نمونه پیش‌بینی FCN8s (ردیف اول)، شبکه Dilation (ردیف دوم)، U-Net (ردیف سوم) و MS-Dilation Net (ردیف چهارم).

منظم‌سازی مکانی - زمانی برای تقسیم‌بندی معنایی

برای منظم‌سازی مکانی و زمانی، از روش‌های مورد استفاده در بهینه‌سازی فضای ویژگی (FSO) استفاده می‌کنیم [۱۱]. همان‌طور که FSO یک بلوک از تصاویر را به‌طور هم‌زمان پردازش می‌کند، یک بلوک از ۵ فریم متوالی بافاصله ۱۰ فریم از فایل‌های ویدئویی ارائه‌شده استخراج می‌شود و تصویر آزمایشی در مرکز هر بلوک قرار می‌گیرد. شکاف بین فریم‌های متوالی خیلی زیاد نیست تا بتوان جریان مناسب را استخراج کرد. امتیاز IoU برای پردازش پس از FSO با Unary map از شبکه‌های مختلف تنظیم‌شده در جدول ۱ گزارش شده است. برای هر مدل، حدود ۱٪ بهبود امتیاز IOU برای هر کلاس جداگانه وجود دارد، به‌جز برای کلاس‌های انسان و ماشین‌های متحرک. FSO بیشتر برای کلاسی که نمونه آن پیکسل‌های تصویر بیشتری را پوشش می‌دهد. امتیاز IOU کلاس انسانی برای شبکه MS-Dilation با حاشیه زیادی، نزدیک به ۴٪ کاهش می‌یابد. نمونه‌ای از پالایش در شکل ۴ نشان داده شده است.

جدول ۲: IOU برای استراتژی‌های آموزشی مختلف امتیاز می‌گیرد. نمرات IOU به صورت درصد گزارش شده و بهترین نتایج به صورت پررنگ نشان داده شده است. w مخفف with و w/o مخفف without است

Model	Buliding	Tree	Clutter	Rroad	Low Vegetation	Static Car	Moving Car	Human	Mean IOH
Fine-Tune W/O AuXiliary Loss	97.5	94.2	94	95.3	91.5	95.4	92.5	91.2	93.75
Fine-Tune W AuXiliary Loss	98.2	94.5	94.8	96.6	92.3	95	93.8	93.4	94.82
Fine-Tune W+W/O AuXiliary Loss	98.4	95.1	93.7	97.7	93.3	98.3	95.8	96.3	96.75



شکل ۴: نمونه‌هایی از منظم‌سازی مکانی و زمانی برای تقسیم‌بندی معنایی شناسه تصویر پهپاد (تصویر سمت چپ تصویر اصلی ما است تصویر وسط پیش‌بینی را بدون اصلاح FSO به علاوه D CRF^۳ نشان می‌دهد. ستون سمت راست پیش‌بینی تصفیه‌شده مربوطه را با پالایش FSO به علاوه D CRF^۳ نشان می‌دهد).

۶- مقایسه و ارزیابی نتایج

در بحث مقایسه با دیگر کارهای صورت گرفت در تقسیم‌بندی معنایی تصاویر پهپاد در مقاله تجزیه و تحلیل عملکرد الگوریتم‌های تقسیم‌بندی معنایی برای مجموعه داده‌های ویدئویی هوایی پهپاد جدید با حاشیه‌نویسی دقیق (ManipalUAVid) برای تقسیم‌بندی معنایی بادقت پیکسل (0.90)، (0.93، CRF)، (0.92، U-Net)، (FCN) و (۰،۹۱)، (DeepLabV3C) دست‌یافته‌اند همچنین در مقاله تقسیم‌بندی معنایی پیشرفته ویدئوهای هوایی پهپاد با جاسازی اطلاعات زمانی برای تقسیم‌بندی معنایی با استفاده از مدل پیشنهادی معماری CNN مبتنی بر رمزگذار- رمزگشا به دقت ۰،۷۹ درصد دست‌یافته‌اند که مدل ترکیبی پیشنهادی ما از نظر دقت نسبت به کار دیگران میانگین ۹۶،۷۵ داشته است که خروجی هر کلاس در جدول ۱ منتشر شده است.

۷- نتیجه‌گیری

در این مقاله، ما یک شناسه مجموعه داده‌های پهپاد جدید برای پیشبرد توسعه بخش‌بندی معنایی در صحنه‌های خیابان شهری از تصاویر پهپاد ارائه کرده‌ایم. مجموعه داده ما چندین چالش را برای کار تقسیم‌بندی معنایی به همراه داشته است، از جمله تنوع در مقیاس بزرگ برای اشیاء مختلف، تشخیص شی متحرک در صحنه‌های خیابان، و سازگاری زمانی در فریم‌های متعدد. هشت کلاس برای وظیفه برچسب‌گذاری معنایی تعریف و برچسب‌گذاری شده است. قابلیت استفاده از شناسه مجموعه داده‌های پهپاد ما نیز با چندین شبکه عصبی کانولوشنال عمیق ثابت شده است. در میان آن‌ها شبکه اتساع چند مقیاسی پیشنهادی بهترین عملکرد را از طریق استخراج ویژگی چند مقیاسی دارد. همچنین نشان داده شده است که پیش آموزش شبکه و اعمال منظم سازی مکانی - زمانی برای کار برچسب‌گذاری معنایی شناسه پهپاد مفید است. اگرچه شناسه مجموعه داده‌های پهپاد دارای محدودیت‌هایی در اندازه و تعداد کلاس‌ها در مقایسه با بزرگ‌ترین مجموعه داده در جامعه بینایی کامپیوتر است، مجموعه داده شناسه UAV در حال حاضر می‌تواند برای مقاصد معیار استفاده شود. در آینده، می‌خواهیم مجموعه داده‌ها را از نظر اندازه و تعداد دسته‌ها بیشتر گسترش دهیم تا برای پیشبرد تحقیقات تقسیم‌بندی معنایی برای تصاویر پهپاد چالش برانگیزتر و مفیدتر شود.

منابع:

- [1] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In CVPR, 2016.
- [2] F. Rottensteiner, G. Sohn, M. Gerke, J.D. Wegner, U. Breitkopf, and J. Jung. Results of the isprs benchmark on urban object detection and 3d building reconstruction. ISPRS journal of photogrammetry and remote sensing, 93:256–271, 2014.
- [3] Gabriel J. Brostow, Jamie Shotton, Julien Fauqueur, and Roberto Cipolla. Segmentation and recognition using structure from motion point clouds. In ECCV, pages 44–57, 2008.
- [4] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In CVPR, June 2015.
- [5] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. In ICLR, 2016.
- [6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In MICCAI, pages 234–241, 2015.
- [7] Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Cocomp: Thing and stuff classes in context. In CVPR, June 2018.
- [8] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Susstrunk, et al. Slic superpix-els compared to state-of-the-art superpixel methods. PAMI, 34(11):2274–2282, 2012.
- [9] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes challenge: A retrospective. IJCV, 111(1):98–136, Jan. 2015.
- [10] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. Tensorflow: a system for large-scale machine learning. In OSDI, volume 16, pages 265–283, 2016.
- [11] Abhijit Kundu, Vibhav Vineet, and Vladlen Koltun. Feature space optimization for semantic video segmentation. In CVPR, pages 3168–3175, 2016.

Semantic segmentation for UAV images using Deep learning methods

Seyyed Mohammad ali Ziyai

In this article, the semantic segmentation of UAV images is discussed in order to identify the objects in the images. Complementary high-resolution UAV semantic segmentation datasets bring new challenges, including large-scale variability, moving object detection, and maintaining temporal consistency. The presented dataset focuses on increasing flight safety and object detection in urban areas. Several deep learning based methods have been presented on datasets. Among these methods, Multi-Scale-Dilation is proposed by extracting multi-scale features, which has the best performance. Its average score reaches 96.75%, which is more than 6.1% better than others. We have also investigated the effect of spatial and temporal regularization for a sequence of data using feature space optimization.

Keywords: semantic segmentation of images, drone datasets, drone identification, drone images, deep learning