مبانی بینایی کامپیوتر

نیم‌سال دوم ۱۴۰۳ - ۱۴۰۲

مدرس: دکتر احمد محمودی

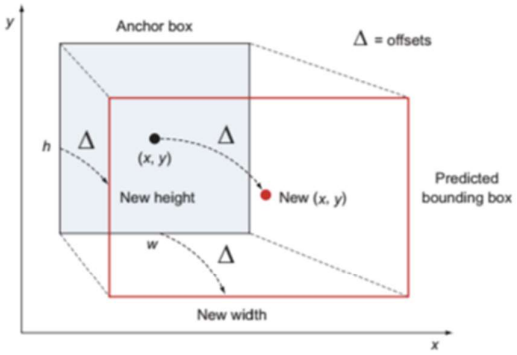
دانشکده علوم کامپیوتر

تمرین سری سوم

* ۱۰ نمره از ۱۱۰ نمره تمرین امتیازی است *

مسئله ۱. (۴)

توابع خطای Fast-RCNN و Faster-RCNN را موشکافانه مورد بررسی قرار دهید. همچنین ضمن بیان ارتباط مفهوم RPN با تصویر زیر[1]، مشخص کنید برای کدام یک مطرح شد و به چه چیزی اشاره دارد و موجب چه پیشرفتی شده است؟



مسئله ۲. (۴)

با درنظرگرفتن شارنوری و الگوریتم‌های Lucas & Kanade و Horn & Schunck به موارد زیر پاسخ دهید.

الف) تفاوت اصلی بین دو الگوریتم را توضیح دهید. ضمن بیان دلیل مقاومت در برابر نویز در روش Lucas & Kanade نسبت به Horn & Schunck ؛ پیش‌فرض های هر یک را نیز با یکدیگر مقایسه کنید.

ب) نقش پارامتر رگولاریزاسیون α در روش Horn-Schunck چیست؟ تغییر آن چگونه بر میدان شار نوری حاصل تاثیر می‌گذارد؟ (ارائه مثال توصیه می‌شود)

ج) مشکل aperture در optical flow و چگونگی تاثیر آن بر motion estimation را توضیح دهید

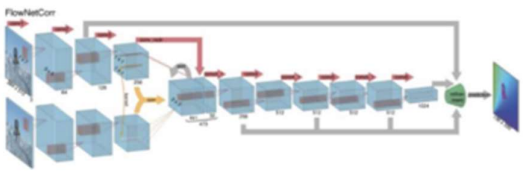
[1] From course slides

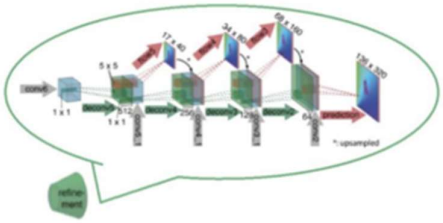
مسئله۳. (۶) ❋

با درنظرگرفتن مدل Flownet به موارد زیر پاسخ دهید.

الف) این مدل دو واریانت اصلی دارد؛ FlownetS و FlownetC. این دو نوع را با تمرکز بر تفاوت های معماری و مزایای خاصی که هر نوع برای تخمین شارنوری ارائه می‌دهند، مقایسه کنید.

ب) اجزای اصلی معماری FlownetC[2] مانند ساختار encoder، decoder، warping operations و لایه های کوریلیشنی و ارتباط این اجزا را مورد بررسی قرار دهید. در مورد چالش های مطرح شده توسط این مدل در تخمین شار نوری، به ویژه برای جابجایی های بزرگ و مناطق مسدود بحث کنید.





مسئله۴.(۶)

ضمن توضیح کامل و با جزئیات تفسیر مفهوم خودتوجهی[3] در زمینه ترنسفرمرهای بینایی؛ یک گونه از آن را با مشخصات زیر درنظر بگیرید و با توجه به آن تعداد پارامتری قابل آموزش مدل و تعداد پچ هایی که از تصویر ورودی تولید می‌شوند را بدست آورید. همچنین مزایا و محدودیت‌های ترنسفرهای بینایی را نسبت به شبکه‌های عصبی کانولوشنی (CNNs) در وظایف بینایی بررسی کنید.

- اندازه تصویر ورودی: ۲۲۴ در ۲۲۴ پیکسل
- اندازه پچ: ۱۶ در ۱۶ پیکسل
- تعداد سرهای توجه: ۸
- بعد تعبیه: ۵۱۲
- تعداد لایه‌ها: ۱۲

[2] From course slides

[3] Self-attention

سوالات کامپیوتری (۹۰)

هدف از انجام این تمرین، بکارگیری روش‌های تشخیص اشیا، تحلیل حرکت و مقدمه ای بر مبدل های بینایی[4] و همچنین پوشش مطالب درس تا انتهای اسلاید ۷ است. لطفاً برای اینکه قالب تمامی پاسخ‌ها یکدست باشد، مراحل زیر را به ترتیب دنبال کنید.

(1) یک دایرکتوری با عنوان CV-CHW3-[Student ID] بسازید.
(2) برای هر سوال، یک نوتبوک با عنوان Q[number].ipynb در آن ایجاد کنید.
(3) برای سوالات موردی ابتدای هر مورد، سلولی مارک‌داون ایجاد کنید.
(4) مورد سوال را در قالب Q[number]-[part] در ابتدای این سلول بنویسید. (مثلا Q4-A)
(5) هر جا سوالات نیاز به پاسخ تئوری داشت؛ مارک‌داون ایجاد کنید و توضیحات لازم را بنویسید.
(6) همگی ضمایم بصورت یکجا در این لینک[5] قابل دسترسی هستند.

مسئله اول کامپیوتری (۱۵)

در این سوال، هدف ما طراحی یک مدل تشخیص شئ ساده برای شناسایی گربه‌ها در تصاویر با الهام از معماری‌های تدریس‌شده در کلاس؛ با طی کردن مراحل ذکر شده و کامل کردن نوتبوک پیوست شده است. در ابتدا داده‌ها را پیش‌پردازش کنید. (به این صورت که تصاویر و حاشیه‌نگاری[6]ها، را بخوانید، ابعاد تصاویر را به اندازه مشخصی (۴۱۶ در ۴۱۶ پیکسل) تغییر دهید و نرمال‌سازی مقادیر پیکسل‌ها و مختصات جعبه‌های محدودکننده[7] است.) در ادامه، معماری مدل را با ساخت یک شبکه هرمی ویژگی[8] برای استخراج ویژگی‌ها از مراحل مختلف ResNet50 به همراه یک متد Detection Head شامل دو شاخه برای رگرسیون جعبه محدود کننده[9] و پیش‌بینی کلاس طراحی کنید. سپس، با یکپارچه‌سازی شبکه پایه[10]، شبکه هرمی ویژگی و detection head، مدل نهایی را بسازید. در پایان، مدل را با استفاده از مجموعه داده آموزش دهید و دقت مدل و نمونه نتایج تشخیص را نمایش دهید. (تمامی ضمایم در Q1.zip قابل مشاهده است)

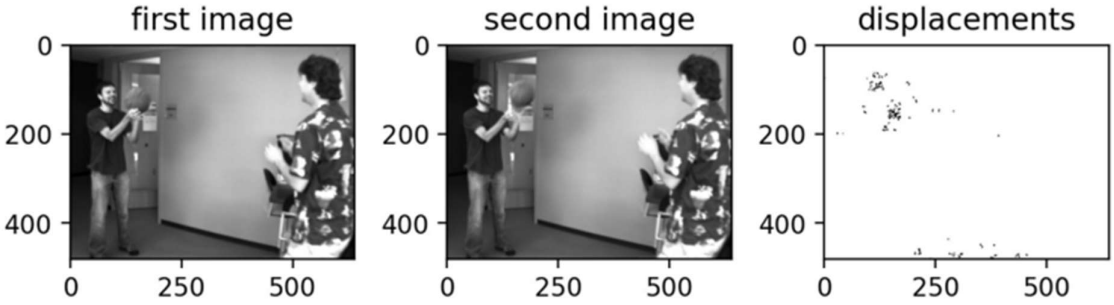
---

4 Vision Transformers
5 https://drive.google.com/drive/folders/1aTEUFJJ43ZRwY1Pw1qs4bCp073KC0n6l?usp=sharing
6 Annotation
7 Bounding box
8 FPN
9 Regression bounding box
10 Backbone

مسئله دوم کامپیوتری (۱۰)

در این مسئله، روش Lucas-Kanade برای تخمین شار نوری و تحلیل حرکت پیاده سازی خواهد شد. شار نوری حرکت اجسام بین فریم های متوالی است. از آنجایی که هر ویدیو دنباله ای از تصاویر است هر دو فریم متوالی را می توان به الگوریتم های محاسبه‌گر شارنوری داد که جابجایی[11] ویژگی ها را به صورت مکانی محاسبه می کند. درنظر داشته باشید فرض شدهدو فریم متوالی گرفته شده از یک ویدیو با اختلاف زمانی ناچیز گرفته شده اند، پس می توان فرض کرد که جابه‌جایی هیچ جسمی زیاد نیست و روشنایی جسم تغییر چشم‌گیری نمی‌کند و همچنین تمام پیکسل های همسایه (شامل شی) در یک جهت حرکت می‌کنند. حال با استفاده از این مفروضات حرکت اجسام را در فریم های داده شده در تصاویر موجود در ضمایم محاسبه کنید.

(استفاده از کتابخانه های آماده برای Lucas-Kanade مجاز نیست و هدف پیاده سازی گام به گام تمام مراحل است.)



(تمامی ضمایم در Q2.zip قابل مشاهده است)

مسئله سوم کامپیوتری (۳۰)

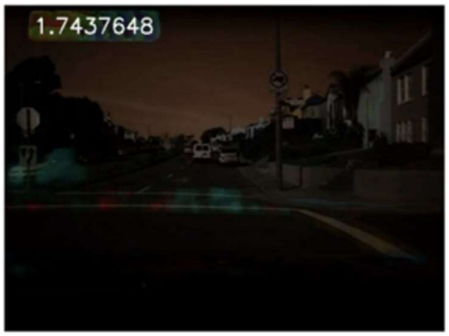
می‌خواهیم با استفاده از داده های دوربین ثبت وقایع یک خودرو درحال حرکت (رو به جلو)، سرعت آن را پیش‌بینی کنیم. این کار به علت شرایط نوری مختلف محیط چالش برانگیز است. یک رویکرد معمول استفاده از CNN ها با تصاویر و داده های برچسب‌گذاری شده است. با این حال، اگر ما از فریم‌های اصلی ویدئو به تنهایی بهره ببریم؛ مدل احتمالا دچار بیش‌برازش خواهد شد. زیرا تنها تصاویر را با مقادیر سرعت به خاطر خواهد سپرد. بنابراین، باید از ویژگی های ورودی مناسب برای آموزش بهره ببریم. جریان نوری با استفاده از دو فریم متوالی ویدیویی در حالت خاکستری، ماتریسی با همان ابعاد تصویر ورودی ایجاد می‌کند. هر پیکسل این ماتریس نشان‌دهنده تغییر موقعیت و سرعت آن پیکسل نسبت به فریم قبلی است. جهت و بزرگی تغییرات در هر پیکسل از طریق کانال رنگی HSV ذخیره می‌شود. با آموزش شبکه با استفاده از جریان نوری، شبکه می‌آموزد که کدام

[11] Displacements

پیکسل‌های جریان نوری مهم هستند و وزن‌دهی مناسب را به آنها اختصاص می‌دهد. حال با استفاده از چهارچوب کاری PyTorch و معماری دلخواه، شبکه ای با استفاده از شار نوری آموزش دهید که سرعت حرکت خودرو را تخمین بزند و با معیار های مناسب دقت مدل خود را گزارش دهید. همچنین در خروجی دو ویدئو یکی با استفاده از ویدئوی اصلی و دیگری ویدئو ترکیبی جریان نوری و ویدئوی تست، هر دو با نمایش سرعت پیش‌بینی شده در بالای تصویر، تولید کرده و بصری سازی کنید.(برای درک بهتر دو فریم از موارد خواسته شده آورده شده است.)

(ملاک ارزیابی این سوال بر روی دادگان آزمایش داده شده و همچنین رعایت موارد خواسته شده است.)

(تمامی ضمایم در Q3.zip قابل مشاهده است)



مسئله چهارم کامپیوتری (۳۵)

در این مسئله، هدف ما استفاده از ترنسفرمر بینایی برای دسته‌بندی تصاویر مجموعه‌داده CIFAR-10 است. ابتدا مجموعه‌داده CIFAR-10 را با تغییر اندازه تصاویر به ۲۲۴ در ۲۲۴پیکسل و نرمال‌سازی مقادیر پیکسل‌ها پیش‌پردازش کنید. به دلیل سنگین بودن عملیات، ترجیحاً این نوت‌بوک را در گوگل کولب بارگذاری کرده و از مسیر زیر استفاده کنید تا شتاب‌دهنده سخت‌افزاری را بر روی GPU یا TPU قرار دهید و از سرعت پردازش بالاتری بهره‌مند شوید.

```
Edit ⇒ Notebook settings ⇒ hardware accelerator
```

در مرحله بعد، یک مدل ترنسفرمر بینایی با استفاده از چارچوب PyTorch پیاده‌سازی کنید. در فایل نوت‌بوک VisionTransformer که در ضمایم این سوال دردسترس است، تمامی مراحل با توضیحات مستندسازی شده در قابل مشاهده هستند و مقادیر هایپرپارامترها از پیش داده شده‌اند. وظیفه شما تکمیل ماژول‌های مشخص شده است تا مدل VisionTransformer بتواند به درستی عمل کند. حال برای این کار ترنسفرمر بینایی را روی

مجموعه‌داده پیش‌پردازش شده CIFAR-10 برای تعداد مشخصی مرحله آموزش دهید. خطا و روند بهینه‌سازی از قبل تعریف شده‌اند، و شما باید حلقه‌ی آموزش را پیاده‌سازی کنید. برای بهبود عملکرد ترنسفرمر بینایی می‌توانید هایپرپارامترهای مختلف را امتحان کنید. بدین منظور، در فایل نوت‌بوک یک سلول جدید ایجاد کرده و هایپرپارامترهای جدید را تعریف کنید و مدل را بر اساس آن ایجاد و حلقه‌ی آموزش را از ابتدا آغاز کنید. بهترین دقت نیز شامل نمره‌ی اضافی خواهد شد. پس از آموزش مدل، عملکرد آن را با محاسبه دقت روی یک مجموعه آزمایش جداگانه ارزیابی کنید. در نهایت، عملکرد ترنسفرمر بینایی را با مدل‌های پایه‌ای مانند ResNet یا VGG بر روی همان مجموعه‌داده مقایسه کنید. برای این منظور، مدل‌های از پیش آموزش‌داده شده موجود در کتابخانه PyTorch را در نوت‌بوک بارگذاری کرده و نتایج آن‌ها را نیز گزارش دهید. (تمامی ضمایم در Q4.zip قابل مشاهده اس

نکات:

- تحویل تکلیف در سامانه کوئرا و تا زمان مشخص‌شده خواهد بود.
- تمارین تایپ شده شامل ۱۵ درصد نمره امتیازی(سوالات تئوری) می‌باشد.
- مسئله سوم تئوری(الف)، چهارم کامپیوتری حاوی مواردی امتیازی (*) هستند
- استفاده از Mathcha توصیه می‌شود.
- فرمت فایلِ سوالات خود را حتماً به صورت زیر رعایت فرمایید.

HW3/CHW3 - [Full Name] - [Student ID]

- در صورت مشاهده هرگونه تقلب، رونویسی و ... با افراد خاطی برخورد خواهد شد.