



## تمرین سری 7 درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدرضا محمدی  
دستیار آموزشی مرتبط: پارسا عیسی زاده –  
شایان موسوی نیا – مهدی خورشیا  
مهلت تحویل: ۱۴۰۲/ تیر / ۱۶

1 در این سؤال قرار است برای مجموعه داده‌ای که در اختیار دارید مدل yolov7 برای تشخیص پنل‌های خورشیدی train کنید. اما این مجموعه داده دارای برچسب‌های مناسب برای آموزش مدل yolov7 نیست و باید برای آن ساخته شود. (60 نمره)

- درون نوت‌بوکی که به پیوست داده شده است، ابتدا از mask‌های هر تصویر استفاده کنید تا کانتورهای پنل‌های خورشیدی را استخراج کنید.
- سپس مستطیل محاط به هر کانتور را به دست آورید.
- از اطلاعاتی که مستطیل محاط شده به دست می‌دهد استفاده کنید تا بتوانید برچسب‌های مناسب هر تصویر برای مدل yolov7 را تولید و ذخیره کنید. برای این کار در نظر داشته باشد مدل yolo انتظار دارد در کنار هر فایل تصویر آموزشی، یک فایل txt. به همان اسم به عنوان برچسب ببیند که درون آن به ازای هر شی درون تصویر یک سطر با اطلاعات زیر وجود داشته باشد:  
0 0.613177 0.452296 0.101875 0.071815  
Class\_id x\_center\_normalized y\_center\_normalized w\_rect\_normalized h\_rect\_normalized  
این اطلاعات با یک space از هم جدا شده‌اند.
- کلاس‌ها از 0 شروع می‌شود. فرض کنید اگر در تصاویرتان 5 کلاس شی دارید، آنگاه idهای شما از 0 تا 4 خواهند بود.
- نحوه normalize کردن مختصات x و y به این صورت است که کافی است مختصات را به تناسب به عرض و طول تصویر تقسیم کنید.
- سپس کد منبع مدل yolov7 را از آدرسی که داده شده است clone کرده و بر اساس گام‌هایی که درون readme آن آورده شده است، مدل را بر روی داده‌های خودآموزش دهید. لطفاً حین آموزش مدل، به تنظیم درست hyperparameterها دقت کنید.

لینک دریافت تصاویر:

[https://drive.google.com/file/d/1KWnX3eMPJrzhsgei0LmyGUbUV5pqKw\\_R/view](https://drive.google.com/file/d/1KWnX3eMPJrzhsgei0LmyGUbUV5pqKw_R/view)

همون فایل تمرین ششم HW6\_q5

2 - همان طور که می‌دانید الگوریتم SORT از سه مرحله اصلی تشکیل شده است:

1. پیدا کردن objectها در frame اولی توسط شبکه Faster R-CNN. الگوریتم Faster R-CNN، آبجکت‌ها را در frame اولیه به تعدادی گره، embed می‌کند.
2. فیلتر Kalman حرکت بعدی object را توسط اعمال ریاضی (بدون استفاده از شبکه deep) پیش‌بینی می‌کند.
3. متناظر کردن objectهای جدید به قبلی‌ها (الگوریتم Hungarian)

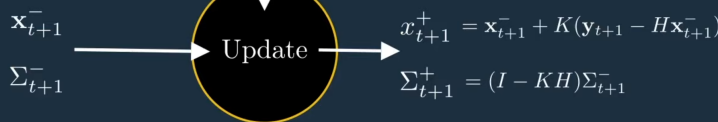
برای توضیح ریاضی فیلتر Kalman می‌توانید از ویدئوی لینک زیر کمک بگیرید:

<https://www.youtube.com/watch?v=IFeCIbljreY>

$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{u}, \quad \mathbf{u} \sim \mathcal{N}(0, P)$  Past Observations

$\mathbf{y}_t = H\mathbf{x}_t + \mathbf{v}, \quad \mathbf{v} \sim \mathcal{N}(0, Q)$  Current Measurements

$$\begin{bmatrix} x_{t+1} \\ \hat{x}_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ \hat{x}_t \end{bmatrix}$$
$$\Sigma_0 = \begin{bmatrix} \Sigma_{xx} & \Sigma_{x\dot{x}} \\ \Sigma_{x\dot{x}} & \Sigma_{\dot{x}\dot{x}} \end{bmatrix}$$



نام مدرس: دکتر محمد

دستیار آموزشی مرتبط: پارسا عیس

شایان موسوی نیا - مه

مهلت تحویل: ۱۴۰۲/

همچنین از source code های زیر می توانید ایده بگیرید:

الگوریتم SORT :

<https://github.com/abewley/sort>

الگوریتم فیلتر Kalman :

[https://github.com/rllabbe/filterpy/blob/master/filterpy/kalman/kalman\\_filter.py](https://github.com/rllabbe/filterpy/blob/master/filterpy/kalman/kalman_filter.py)

حال به سؤالات زیر پاسخ دهید (20 نمره):

- اعمال ریاضی فیلتر Kalman را برای پیش بینی 1 پارامتر توضیح دهید.  $w, h, x, y$
- اعمال ریاضی Kalman را به شکل ماتریسی برای پیش بینی حرکت labelها (با 4 ویژگی) در شبکه‌ای که عکس‌ها را به 7 گره embed می‌کند توضیح دهید.
- تفاوت الگوریتم Deep SORT با الگوریتم SORT (همان‌طور که از اسمش معلوم است) این است که از یک شبکه Deep استفاده می‌کند. به چه منظوری از شبکه Deep استفاده می‌شود؟

3 - الف) یکی از مهم‌ترین بخش‌های ساخت یک مدل هوش مصنوعی، فراهم کردن مجموعه داده‌های آموزشی است. در این سؤال، هدف ما برچسب‌زنی یک سری فریم از یک قفس است که در آن تعدادی پرند قرار دارند که با آن بتوانیم داده‌های آموزش مورد نیاز خود را بسازیم. برای برچسب‌زنی، در این سؤال قصد داریم از نرم‌افزار cvat (<https://www.cvat.ai/>) استفاده کنیم. پس از اتمام برچسب‌زنی، خروجی برچسب‌ها را به فرمت yolov5 از cvat استخراج می‌کنیم و با استفاده از آن، مدل yolov5s را آموزش می‌دهیم. در این آموزش، batch size برابر با 2 و تعداد epochها برابر با 50 خواهد بود. نتایج آموزش و برچسب‌های زده شده را در پاسخ خود ارسال کنید. (فایل‌های مورد نیاز این سؤال در لینک زیر موجود است) (40 نمره)

<https://drive.google.com/drive/folders/1m004UAo-h7A5Vt5eTqcrvEYANh9iQ0G4?usp=sharing>

3 - ب) در بخش بالا دیدیم که چگونه با داشتن مجموع داده مناسب، می‌تواند یک مدل هوش را برای object detection آموزش داد. آیا با آموزش مدل بر روی یک مجموعه داده که پرندگان آن در قفس نیستند نیز می‌توان مدلی خوبی ساخت؟ در صورت موافق بودن یا مخالف بودن، دلیل خود را بیان کنید. (10 نمره)

3 - ج) در خیلی از حوزه‌ها، فراهم کردن مجموعه داده برای موارد خاص، سخت خواهد بود و همچنین برچسب‌زنی بر روی یک مجموعه داده خام، بسیار زمان بر است. راهکاری ارائه دهید که بتوان از آن برای detection پرندگان در قفس، با استفاده از مجموعه داده‌های آماده (مانند COCO) استفاده کرد که با آن بتوان مدلی را فراهم کرد که پرندگان را در قفس نیز به خوبی بشناسد. (10 نمره)



## تمرین سری 7 درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمدرضا محمدی  
دستیار آموزشی مرتبط: پارسا عیسی زاده -  
شایان موسوی نیا - مهدی خورشیا

مهلت تحویل: ۱۴۰۲ / تیر / ۱۶

Siamese Fully Convolutional

4 - به سؤالات زیر در مورد **siamFC** پاسخ دهید: (20 نمره)

- مزایا و محدودیت‌های معماری شبکه siamFC در وظایف بینایی کامپیوتر چیست؟
- شبکه سیامی (SiamFC) در ردیابی اشیاء بصری چگونه عمل می‌کند و اجزای اصلی آن چیست؟
- چالش‌های مرتبط با ردیابی شی چیست و SiamFC چگونه آنها را برطرف می‌کند؟
- چگونه مفهوم معماری سیامی فراتر از ردیابی شی گسترش می‌یابد و چه کارهای بینایی کامپیوتری دیگری می‌تواند از آن بهره‌مند شود؟ (به طور مثال، One-Shot Learning یکی از حوزه‌های مورد استفاده این شبکه است).

لطفاً سند قوانین انجام و تحویل تمرین‌های درس را مطالعه و موارد خواسته شده را رعایت کنید.

موفق باشید.

پاسخ دهید siamFC به سؤالات زیر در مورد:

- در وظایف بینایی کامپیوتر چیست؟ siamFC مزایا و محدودیت‌های معماری شبکه
- در ردیابی اشیاء بصری چگونه عمل می‌کند و اجزای اصلی آن چیست؟ (SiamFC) شبکه سیامی
- چگونه آنها را برطرف می‌کند؟ SiamFC چالش‌های مرتبط با ردیابی شی چیست و
- چگونه مفهوم معماری سیامی فراتر از ردیابی شی گسترش می‌یابد و چه کارهای بینایی کامپیوتری دیگری می‌تواند از آن بهره‌مند شود؟ (به یکی از حوزه‌های مورد استفاده این شبکه است One-Shot Learning طور مثال)

الف) یکی از مهم‌ترین بخشهای ساخت یک مدل هوش مصنوعی، فراهم کردن مجموعه داده‌های آموزشی است. در این سؤال، هدف ما برچسب زنی یک سری تصاویر از یک قفس است که در آن تعدادی پرندۀ قرار دارند که با آن بتوانیم داده‌های آموزش مورد نیاز خود را بسازیم. برای برچسب زنی، در این سؤال قصد داریم از استخراج می‌کنیم و با cvat از yolov5 استفاده کنیم. پس از اتمام برچسب زنی، خروجی برچسب‌ها را به فرمت [cvat \(https://www.cvat.ai\)](https://www.cvat.ai) نرم‌افزارها برابر با 50 خواهد بود. نتایج آموزش و برچسب‌های epoch برابر با 2 و تعداد batch size، را آموزش می‌دهیم. در این آموزش yolov5s استفاده از آن مدل (زده شده را در پاسخ خود ارسال کنید). (فایل‌های مورد نیاز این سؤال در لینک زیر موجود است)

<https://drive.google.com/drive/folders/1m004UAo-h7A5Vt5eTqcrvEYANh9iQ0G4>

آموزش داد. آیا با آموزش مد بر روی object detection (ب) در بخش بالا دیدیم که چگونه با داشتن مجموعه داده مناسب می‌توان یک مدل را برای یک مجموعه داده که پرندگان آن در قفس نیستند نیز می‌توان مدلی خوبی ساخت؟ در صورت موافقت یا مخالف بودن، دلیل خود را بیان کنید. (ج) در خیلی از حوزه‌ها، فراهم کردن مجموعه داده برای موارد خاص، سخت خواهد بود و همچنین برچسب زنی بر روی یک مجموعه داده خام (coco) پرندگان در قفس، با استفاده از مجموعه داده‌های آماده (مانند detection بسیار زمان‌بر است. راهکاری ارائه دهید که بتوان از آن برای استفاده کرد که با آن بتوان مدلی را فراهم کرد که پرندگان را در قفس نیز به خوبی بشناسد.