



تمرین سری ششم درس مبانی بینایی کامپیوتر

نام مدرس: دکتر محمد رضا محمدی

دستیاران آموزشی مرتبط: علیرضا خیرخواه، امیرحسین حسینی، کسری شریعی،
آرین ستاری

مهلت تحولی (بدون کسر نمره): سه شنبه 27 خرداد

بخش تئوری

1. در ارتباط با object detection به سوالات زیر پاسخ دهید: (21 نمره)

- الف) برای پیدا کردن پریوپزال‌ها در بعضی مواقع معیار IOU خوب جواب نمیدهد (کیفیت لازم را ندارد)، آن موارد که IOU دچار خطا می‌شود را نام بده و معیارهای مکمل یا جایگزین برای ارزیابی بهتر را بنویسید.

مقاله پیشنهادی: [Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression]

ب) Proposals with high recall but at low overlap are not effective for detection. این جمله را تحلیل کرده و با مثال توضیح دهید.

مقاله پیشنهادی: [What makes for effective detection proposals?]

پ) تفاوت anchor boxes در SSD با سایر مدل‌های detection مانند Faster R-CNN چیست؟

ت) چرا در SSD نیازی به استفاده از Selective Search region proposal methods مانند RPN وجود ندارد؟

ث) در مقاله SSD، از hard negative mining استفاده شده است. این تکنیک چیست و چرا برای آموزش پایدار شبکه مهم است؟

مقاله پیشنهادی: [SSD: Single Shot MultiBox Detector]

ج) در مقاله "SSD"، نویسنده‌گان عملکرد SSD را از نظر میانگین دقت (mAP) و نرخ فریم بر ثانیه (FPS) با مدل‌های Faster R-CNN و YOLOv1 مقایسه کرده‌اند. با استناد به اعداد و نتایج ارائه شده در مقاله، مشخص کنید:

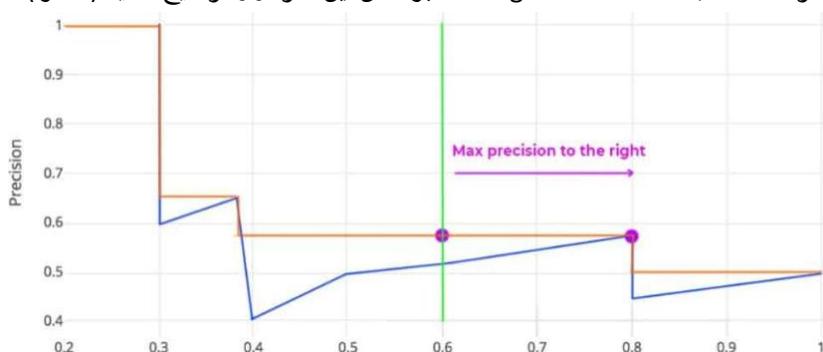
SSD از نظر دقت و سرعت در چه جایگاهی قرار دارد؟ •

چه (trade-off) ای میان سرعت و دقت ارائه می‌دهد؟ •

برای چه کاربردهایی SSD نسبت به سایر روش‌ها انتخاب بهتری است؟ •

مقاله پیشنهادی: [SSD: Single Shot MultiBox Detector]

2. در نمودار مختصاتی زیر که محور افقی نشان‌دهنده Recall و محور عمودی Precision است، دو منحنی رسم شده‌اند. یکی از این منحنی‌ها نمایانگر مقادیر واقعی (Actual Precision) و دیگری نمایانگر Precision درونیابی شده (Interpolated Precision) است که در محاسبه‌ی Average Precision به کار می‌رود. مشخص کنید کدام منحنی actual و کدام interpolated است، و علت انتخاب خود را تحلیل کنید. چرا Interpolated Precision لازم است؟ هر کدام چه کاربردی را پوشش میدهد؟ همچنین، نحوه‌ی محاسبه‌ی Average Precision براساس این نمودار را توضیح دهید. (8 نمره)





تمرین سری ششم درس مبانی بینایی کامپیوتر

مهلت تحولی (بدون کسر نمره): سه شنبه 27 خرداد

3. سوالات زیر در ارتباط با Semantic Segmentation پاسخ دهید (12 نمره)

(الف) منظور از FCN در چیست؟ چه تفاوتی با معماری های CNN معمولی برای classification دارد؟

چرا در معماری FCN، لایه های Fully Connected حذف شده اند؟ این حذف چه مزایا و چالش هایی ایجاد می کند؟

(ب) از FCN از pre-trained classification models مثل AlexNet, VGG, GoogLeNet استفاده می کند.

چرا استفاده از segmentation در transfer learning کاربردی است؟

(ج) چرا CNN ها باعث از دست رفتن دقیق شود و FCN برای حل آن چه راه حلی پیشنهاد می کند؟

4. به سوالات زیر در ارتباط با Focal Loss پاسخ دهید (11 نمره)

(الف) هنگام استفاده از Focal Loss برای مقابله با class imbalance، این تکنیک چه تاثیری بر precision و recall دارد؟ از نظر

مفهومی آیا میتوان گفت focal loss نسخه‌ی ریگولاریزه شده‌ی cross entropy است؟ اگر γ از صفر تا یک نهایت تغییر کند،

چه اتفاق می افتد؟ دقیق تحلیل کنید.

(ب) اگر training از Focal Loss حذف شود و به جای آن از Cross-Entropy معمولی استفاده شود، چه پدیده ای در دیده میشود؟

(ج) فرض کنید بخواهید RetinaNet را روی یک داده با 100 کلاس آموزش دهید که تنها 10 کلاس آن متداول اند. چه تغییری در focal loss یا تنظیمات anchor پیشنهاد می دهد؟

5. در الگوریتم (SORT) (Simple Online and Realtime Tracking)، از الگوریتم Hungarian برای به حداقل رساندن هزینه کلی

تخصیص n تشخیص (detections) به m مسیر ردیابی (tracks) استفاده می شود. این تخصیص بر اساس یک ماتریس هزینه $C \in \mathbb{R}^{n \times n}$ صورت می گیرد.

(الف) با توجه به ماتریس هزینه زیر، تخصیص بهینه را که هزینه کل را حداقل می کند، پیدا کنید. (17 نمره)

$$C = \begin{bmatrix} 4 & 1 & 3 \\ 2 & 0 & 5 \\ 3 & 2 & 2 \end{bmatrix}$$

(ب) چرا ماتریس هزینه در عمل باید هم فاصله IoU و هم خطای پیش‌بینی حرکت را ترکیب کند؟

(پ) شبکه سیاعی را در نظر بگیرید که تصاویر ورودی را به یک بردار بازنمایی (Embedding) در فضای \mathbb{R}^d نگاشت می کند و از

یکتابع فاصله $d(x_i, x_j) = \|f(x_i) - f(x_j)\|_2^2$ استفاده می کند.

اولاً) نشان دهید که اگر $f(x)$ یکتابع خطی باشد یعنی $f(x) = Wx$ آنگاه تابع فاصله به یک فرم درجه دو تبدیل می شود.

ثانیاً) فرض کنید امتیاز شباهت $S(x_i, x_j) = \exp(-\gamma d(x_i, x_j))$ باشد. وقتی $\gamma \rightarrow \infty$ چه اتفاق برای این امتیاز می افتد؟

این چه معنایی برای اطمینان مدل دارد؟

 دانشگاه علم و صنعت ایران	تمرين سري ششم درس مبانی يينياني كمپيوتر	نام مدرس: دکتر محمد رضا محمدی دستیاران آموزشی مرتبط: علیرضا خیرخواه، امیرحسین حسینی، کسری شریعی، آرین ستاری مهلت تحويل (بدون كسر نمره): سه شنبه 27 خداد
---	--	--

6. ساختار زير يك شبکه‌ی ساده‌ی کانولوشنی برای طبقه‌بندی تصویر با اندازه‌ی ثابت 32×32 است. این شبکه شامل يك لایه‌ی کانولوشنی و دو لایه‌ی Fully Connected است: (18 نمره)

```
model = tf.keras.Sequential()
model.add(tf.keras.layers.Conv2D(16, (3, 3), activation='relu', input_shape=(32, 32, 3)))
model.add(tf.keras.layers.Flatten())
model.add(tf.keras.layers.Dense(128, activation='relu'))
model.add(tf.keras.layers.Dense(10, activation='softmax'))
model
```

(الف) کد همارز اين شبکه را به صورت يك شبکه‌ی تمام‌کانولوشنی بنويسيد به‌گونه‌اي که از لایه‌های Dense و Flatten استفاده نشود و رفتار کلي مدل حفظ شود.

(ب) تعداد کل پارامترهای مدل را در هر دو نسخه (اصلی و تمام‌کانولوشنی) محاسبه و با هم مقایسه کنيد.

(ج) با وجود برابر بودن تعداد پارامترها، دست کم دو دليل برای استفاده از شبکه‌های تمام‌کانولوشنی در مسائل semantic segmentation بيان کنيد.

بخش عملی

. 7. به نوت‌بوک semantic segmentation مراجعه کنید و بخش‌های خواسته شده را کامل کنید. (20 نمره)

همچنین callback‌های مناسب جهت ذخیره بهترین مدل را به کد اضافه کنید. در این سوال ابتدا يك شبکه Unet را از صفر آموزش می‌دهيد.

سپس در بخش دوم يك شبکه Unet با انکودر پیش‌آموخته را آموزش می‌دهيد.

در مرحله اول انکودر کاملاً فریز است و تنها دیکودر آموزش می‌بیند و در مرحله نهایی کل شبکه تنظیم دقیق می‌شود.

در گزارش خود علاوه بر توضیح کد خودتان، توضیحات مختصری در مورد قسمت‌های آماده کد ارائه دهيد.

. 8. دو سوال در دو نوت‌بوک 1_8 و 2_8 در مورد تعريف anchor boxes برای feature map ساده و همچنین اعمال اعمال (NMS (Non-Maximum Suppression))

روی چند جعبه فرضی تهیه و پیوست شده است. (18 نمره)

. 9. دو سوال در نوت‌بوک در مورد ردیابی اشیاء و همچنین تشخیص صورت تهیه و پیوست شده است. (25 نمره)