



دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر



## درس مدل‌های مولد عمیق

مدرس: دکتر مصطفی توسلی‌پور

تمرین شماره ۲

آبان ۱۴۰۴

## فهرست

۳	سوال اول : Normalizing Flow
۳	بخش اول : بخش تئوری (۱۰ نمره)
۴	بخش دوم : بخش عملی
۴	زیر بخش اول : پیاده سازی (۱۰ نمره)
۵	زیر بخش دوم : آموزش مدل و تولید عکس (۱۵ نمره)
۶	زیر بخش سوم : آشکار سازی ناهنجاری (۲۰ نمره)
۷	سوال دوم: CYCLEGAN
۸	بخش اول: سوالات تئوری
۸	زیر بخش اول:تابع زیان مدل (۱۰ نمره)
۹	زیر بخش دوم: معماری و فرایند آموزش مدل (۱۰ نمره)
۹	زیر بخش سوم: محدودیت های مدل (۲ نمره)
۱۰	بخش دوم: پیاده سازی عملی
۱۰	زیر بخش اول: آماده سازی داده ها (۲ نمره)
۱۰	زیر بخش دوم: پیاده سازی معماری مدل (۱۱ نمره)
۱۰	زیر بخش سوم: پیاده سازی تابع هدف و آموزش مدل (۱۵ نمره)
۱۱	زیر بخش چهام: امتیازی (۵ نمره)
۱۱	مراجع
۱۲	نکات تحويل

## سوال اول : NORMALIZING FLOW :

### بخش اول : بخش تئوری

زیر بخش اول: در هر یک از موارد زیر، با استفاده از فرمول change of variable، فرم بسته  $P_X$  را به دست آورید. (۵ نمره)

$$Z \sim U(0, 1), X = 2Z + 1 \quad .1$$

$$Z \sim U(0, 2), X = \exp(Z) \quad .2$$

زیر بخش دوم: در هر یک از تبدیلات زیر، فرم بسته دترمینان ماتریس ژاکوبین را به دست آورید. (۵ نمره)

$$x = (x_1, x_2), z = (z_1, z_2) \text{ s.t. } z_1 = x_1 \text{ and } z_2 = x_2 + m(x_1) \text{ where } m \text{ is a neural network} \quad .1$$

$$x = (x_1, x_2), z = (z_1, z_2) \text{ s.t. } z_1 = x_1 \text{ and } z_2 = s(x_1)x_2 + t(x_1) \text{ where } s \text{ and } t \text{ are neural networks} \quad .2$$

## بخش دوم : بخش عملی

### زیر بخش اول : پیاده سازی (۱۰ نمره)

در این قسمت، از شما انتظار می رود یک Masked Autoregressive Flow<sup>۱</sup> پیاده سازی کنید. همانطور که می دانید، این مدل، یک نوع Normalizing Flow است که در آن هر بلاک تبدیل، یک تبدیل autoregressive است؛ به عبارتی، هر بعد از خروجی تبدیل، تنها وابسته به بعد های قبلی است. برای رسیدن به این مهم، در مقاله MADE<sup>۲</sup> Masked Autoregressive Flow، از یک لایه MADE استفاده شده است.

در ادامه، یک معماری پیشنهادی جهت پیاده سازی این مدل ارائه می شود؛ توجه کنید که لزومی به استفاده از این معماری نیست و در صورت داشتن دسترسی به GPU می توانید از معماری های پیچیده تر استفاده کنید.

● ابعاد تصویر : ۱۲۸ \* ۱۲۸

● معماری MADE

```
Linear(input : 128*128*3, output : 512)
Linear(input:512, output:512)
Linear(input:512, output:2*128*128*3)
```

توجه کنید که لایه Linear، یک لایه خطی معمولی نیست و می بایست شرط autoregressive بودن را حفظ کند.

● تعداد بلاک های تبدیل : ۷

● اندازه دسته<sup>۳</sup> : ۸

<sup>۱</sup> [Masked Autoregressive Flow for Density Estimation](#)

<sup>۲</sup> [MADE: Masked Autoencoder for Distribution Estimation](#)

<sup>۳</sup> Batch size

● تعداد ایپاک : ۱۰۰

● نرخ یادگیری : ۰.۰۰۰۱

در این قسمت، جزئیات مربوط به پیاده سازی MADE و تک بلاک MAF را گزارش کار خود توضیح دهید.

---

### زیر بخش دوم: آموزش مدل و تولید عکس (۱۵ نمره)

پس از پیاده سازی، می بایست مدل خود را روی دیتابست mvtecad، کلاس capsule، عکس های train آموزش دهید. از طریق دستور های زیر می توانید این دیتابست را دانلود کنید.

```
wget
```

```
https://www.mydrive.ch/shares/38536/3830184030e49fe74747669442f0f282/download  
/420937454-1629951595/capsule.tar.xz
```

```
tar -xf capsule.tar.xz
```

برای اطمینان حاصل کردن از عملکرد درست مدل، ۵ عکس تولید کنید. توجه کنید که تولید عکس در این مدل زمانبر است. با استفاده از کتابخانه tqdm در کد خود، می توانید سرعت تولید عکس را مشاهده کنید.

۱. مدت زمان آموزش مدل و همچنین مدت زمان تولید ۵ عکس را مقایسه کنید.

۲. توضیح دهید که چرا در مدل های autoregressive، تولید عکس زمانبر است.

۳. تصاویر تولید شده را در گزارش کار خود بیاورید.

۴. همانطور که مشاهده کردید، MAF در تولید تصاویر کند است. توضیح دهید مدل Inverse Autoregressive Flow<sup>۱</sup> چگونه در تولید تصویر سریع تر است؟ این دو مدل را از لحاظ زمان آموزش مقایسه کنید.

---

<sup>1</sup> Improving Variational Inference with Inverse Autoregressive Flow

---

### زیر بخش سوم: آشکار سازی ناهنجاری<sup>۱</sup> (۲۰ نمره)

در این قسمت، با یکی از کاربردهای مدل‌های Normalizing Flow آشنا می‌شویم. این مساله، است. به طور کلی، مدل‌ها باید تنها با دیدن تصاویر normal، بتوانند anomaly Detection را تشخیص دهند.

۱. در ارزیابی مدل‌های anomaly detection، از معیار دقیقت استفاده نمی‌شود. دلیل آن چیست؟ از

چه معیاری برای ارزیابی این مدل‌ها استفاده می‌شود؟

۲. مفهوم anomaly score را توضیح دهید.

۳. توضیح دهید چگونه می‌توان از یک normalizing flow که روی تصاویر normal یک دیتاست آموزش داده شده، برای تسک anomaly detection استفاده کرد.

حال می‌خواهیم از مدل آموزش داده شده در قسمت قبل، برای Anomaly Detection روی دیتاست evaluate capsule استفاده کنیم. برای این منظور، می‌بایست داده موجود در پوشه test را روی مدل خود کنید. برای هر تصویر، Negative Log Likelihood را به عنوان anomaly score در نظر بگیرید.

۱. معیارهایی که در پرسش ۱ قسمت قبل معرفی کردید را برای مدل خود گزارش کنید. توجه که انتظار می‌رود حداقل به auroc بالای ۶۰٪ برسید.

۲. همانطور که می‌دانید، در مدل‌های VAE، محاسبه دقیق likelihood ممکن نیست؛ اما جالب است بدانید که از این مدل‌ها نیز در anomaly detection استفاده می‌شود؛ بدین صورت که مدل روی تصاویر نرمال آموزش داده می‌شود. سپس در فاز تست، انتظار می‌رود که مدل تصاویر را به صورت normal بازسازی کند؛ به عبارتی با یک دور forward pass از encoder و decoder مدل، تصویر anomaly به صورت نرمال reconstruct می‌شوند. سپس با محاسبه اختلاف بین تصویر اصلی و تصویر reconstruct شده، به anomaly score می‌رسیم. با ذکر دلیل توضیح دهید که آیا از یک normalizing flow نیز می‌توان به این روش در تسک anomaly detection استفاده کرد؟

## سوال دوم: CYCLEGAN

مسئله تبدیل تصویر-به-تصویر<sup>۱</sup>، به دنبال یافتن نگاشتی از یک دامنه تصویری به دامنه‌ای دیگر است، به‌گونه‌ای که محتوای اصلی و ساختار کلی تصویر حفظ و سبک و ویژگی‌های دامنه مقصد در آن القا شود. این تکنیک کاربردهای مهمی دارد که از جمله آن‌ها می‌توان به انتقال سبک هنری (مانند تبدیل عکس به نقاشی)، ترمیم تصاویر (شامل رنگ‌آمیزی تصاویر سیاه‌وسفید و افزایش کیفیت)، پردازش تصاویر پزشکی (مانند تبدیل اسکن‌های MRI به CT برای تشخیص دقیق‌تر)، و بهروزرسانی خودکار نقشه‌ها از طریق تصاویر ماهواره‌ای اشاره کرد.

در این سوال از این تمرین، به بررسی یکی از روش‌های مهم و تأثیرگذار در این زمینه خواهیم پرداخت. ابتدا مقاله [CycleGAN](#) را به دقت مطالعه فرمایید.

---

image-to-image translation <sup>۱</sup>

## بخش اول: سوالات تئوری

### زیر بخش اول: تابع زیان مدل (۱۰ نمره)

۱. روش‌های تبدیل تصویر-به-تصویر مبتنی بر یادگیری نظراتی<sup>۱</sup> مانند ([pix2pix](#))، برای بسیاری از سناریوهای دنیای واقعی با یک چالش اساسی در مرحله جمع‌آوری داده‌های آموزشی مواجه هستند. چرا این چالش به راحتی قابل رفع نیست؟
۲. ایده اصلی CycleGAN برای دور زدن این محدودیت داده‌ای، افزودن یک قید<sup>۲</sup> خاص به فرایند آموزش مدل است. همراه با نوشتمن صورت‌بندی ریاضی، آن را شرح دهید.
۳. چرا تحمیل کردن این قید در این مسئله، حیاتی است؟ به عبارت دیگر، با حذف آن و تکیه صرف به آموزش تخاصمی<sup>۳</sup>، چه رفتار نامطلوبی از مدل مشاهده می‌شود؟
۴. رابطه کامل تابع زیان CycleGAN را بنویسید. در این رابطه،  $\lambda$  چه نقشی دارد؟ اگر مقدار این ضریب، بیش از حد بزرگ، انتخاب شود، چه تاثیری بر خروجی مدل خواهد گذاشت؟
۵. در قسمتی از مقاله، به زیان همانی<sup>۴</sup> اشاره شده است. این زیان چه مفهومی دارد؟ با ارائه یک مثال توضیح دهید که زیان همانی از بروز چه نوع تغییرات ناخواسته‌ای در تصویر جلوگیری می‌کند.

supervised learning <sup>۱</sup>

constraint <sup>۲</sup>

adversarial training <sup>۳</sup>

identity loss <sup>۴</sup>

---

## زیر بخش دوم: معماری و فرایند آموزش مدل (۱۰ نمره)

۱. چرا در CycleGAN، به جای استفاده از یک مولد<sup>۱</sup> واحد برای انجام هر دو جهت تبدیل، از دو مولد مجزا استفاده می‌شود؟ دلایل این انتخاب را از منظر تئوری و عملی بیان کنید.
۲. جزئیات معماری انتخاب شده برای شبکه مولد را با معماری مولد در مدل pix2pix، با توجه به تفاوت در ماهیت وظیفه‌ها، به صورت خلاصه ولی دقیق و کامل مقایسه کنید.
۳. برای شبکه تمیزدهنده<sup>۲</sup> از PatchGAN استفاده شده است. هدف از این انتخاب چیست؟ خروجی این نوع تمیزدهنده را تفسیر کنید.
۴. چرا در آموزش تمیزدهنده، به جای استفاده مستقیم از آخرین تصاویر تولیدشده توسط مولد، از یک بافر تاریخچه تصاویر<sup>۳</sup> استفاده می‌شود؟ این مکانیزم چه مشکلی را در فرآیند آموزش حل می‌کند؟

---

## زیر بخش سوم: محدودیت‌های مدل (۲ نمره)

۱. CycleGAN در وظایفی که نیازمند تغییرات هندسی بزرگ هستند، عملکرد ضعیفی دارد. این محدودیت به چه معناست؟ چرا قید سازگاری چرخه‌ای<sup>۴</sup> مانع از یادگیری این نوع تغییرات می‌شود؟

---

generator <sup>۱</sup>

discriminator <sup>۲</sup>

image history buffer <sup>۳</sup>

cycle consistency <sup>۴</sup>

## بخش دوم: پیاده‌سازی عملی

### زیر بخش اول: آماده‌سازی داده‌ها (۲ نمره)

یکی از مجموعه تصاویر را دریافت نموده و داده‌های آموزش و اعتبارسنجی را آماده کنید. چند تصویر تصادفی از هر دو دامنه (کلاس) نمایش دهید.

می‌توانید به جای مجموعه داده‌های فوق، [monet2photo](#) یا [summer2winter\\_yosemi](#) را نیز انتخاب نمایید.

### زیر بخش دوم: پیاده‌سازی معماری مدل (۱۱ نمره)

ساختار شبکه مولد و تمیزدهنده را در بخش قبلی بررسی کردیم. با توجه به جزئیات معماری‌ها که به صورت شفاف در مقاله ذکر شده است، کلاس‌های Generator و Discriminator را پیاده‌سازی کنید. پیشنهاد می‌شود که ابعاد تصاویر را  $128 \times 128$  در نظر بگیرید.

### زیر بخش سوم: پیاده‌سازیتابع هدف و آموزش مدل (۱۵ نمره)

در بخش قبلی رابطه کامل تابع زیان را دیدیم. آن را پیاده‌سازی کنید و مدل را به اندازه ۲۰ ایپاک آموزش دهید. برای مقداردهی به ابرپارامترها، پیشنهاد می‌شود از مقادیر ذکر شده در مقاله استفاده کنید. در غیر این صورت، دلیل انتخاب مقادیر را بیان کنید. تمام مقادیر ابرپارامترها را حتماً در گزارش ذکر نمایید.

نمودارهای زیان‌ها در طول آموزش را نمایش داده و تحلیل کنید. برای این منظور، حداقل ۵ نمودار (۲ نمودار برای مولدها، ۲ نمودار برای تمیزدهندها، ۱ نمودار برای زیان سازگاری چرخه‌ای) ضروری است. اگر برای تحلیل فرایند آموزش مدل، به نمودارهای دیگری نیز، نیاز است، در گزارش ارائه دهید.

به ازای سه ایپاک ابتدایی، میانی و انتهایی، خروجی مدل را بررسی کنید. به عبارتی دیگر، تعدادی تصویر تصادفی از هر دو دامنه انتخاب کنید و به مدل ورودی دهید. تصاویر تولید شده توسط مدل را به همراه تصاویر ورودی نمایش دهید. نحوه و کیفیت یادگیری مدل را تحلیل کنید.

---

### زیر بخش چهام: امتیازی (۵ نمره)

زیان همانی و بافر تاریخچه تصاویر را پیاده‌سازی کرده و مدل را روی یک مجموعه داده دیگر آموزش دهید.

مشابه زیربخش قبلی، فرایند آموزش مدل و خروجی آن در طول آموزش را تحلیل نمایید.

### مراجع

- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017) (pp. 1125–1134).
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017) (pp. 2223–2232).

## نکات تحویل

- مهلت ارسال این تمرین تا پایان روز "دوشنبه، ۳ آذر" خواهد بود.
- این زمان قابل تمدید نیست و در صورت نیاز می‌توانید از grace time استفاده کنید.
- در نظر داشته باشید که حداقل مهلت آپلود تمرین در سامانه تا ۷ روز پس مهلت تحویل است و پس از آن سامانه بسته خواهد شد.
- پیاده سازی با زبان برنامه نویسی پایتون باید باشد و کدهای شما باید قابل اجرا بوده و به همراه گزارش آپلود شوند.
- انجام این تمرین به صورت یک نفره می‌باشد.
- در صورت مشاهده هر گونه تشابه در گزارش کار یا کدهای پیاده‌سازی، این امر به منزله تقلب برای طرفین در نظر گرفته خواهد شد.
- استفاده از کدهای آماده بدون ذکر منبع و بدون تغییر به منزله تقلب خواهد بود و نمره تمرین شما صفر در نظر گرفته می‌شود
- در صورت رعایت نکردن فرمت گزارش کار نمره گزارش به شما تعلق نخواهد گرفت.
- تحویل تمرین به صورت دستنویس قابل پذیرش نیست.
- تمامی تصاویر و جداول مورد استفاده در گزارش کار باید دارای توضیح (caption) و شماره باشند.
- بخش زیادی از نمره شما مربوط به گزارش کار و روند حل مسئله است.
- لطفاً گزارش، فایل کدها و سایر ضمایم مورد نیاز را با فرمت زیر در سامانه بارگذاری نمایید.
- HW2\_[Lastname]\_[StudentNumber].zip
- در صورت وجود سوال و یا ابهام میتوانید از طریق رایانمeh زیر با موضوع DGM\_HW2 با دستیاران آموزشی در ارتباط باشید:

[javadkavian8@gmail.com](mailto:javadkavian8@gmail.com)

○ سوال اول

[alireza-zamani@outlook.com](mailto:alireza-zamani@outlook.com)

○ سوال دوم

با آرزوی سلامتی و موفقیت روزافزون.