به نام خدا

مبانی بینایی کامپیوتر (دکتر سیفی پور)

دستیار ارشد: پارسا دربان

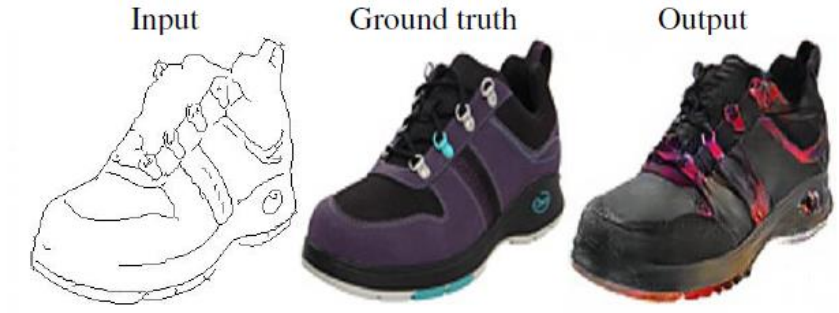
طراح : پارسا دربان

پروژه نهایی

نیم‌سال اول **۱۴۰۵-۱۴۰۴**

مقدمه

در دنیای امروزی، با گسترش روزافزون داده‌های تصویری و پیشرفت چشمگیر روش‌های یادگیری عمیق، پردازش و تحلیل تصاویر به یکی از حوزه‌های کلیدی هوش مصنوعی تبدیل شده است. یکی از شاخه‌های مهم در این زمینه، Image Translation است که هدف آن تبدیل یک تصویر از یک دامنه یا سبک مشخص به دامنه‌ای دیگر، در حالی که ساختار و محتوای اصلی تصویر حفظ می‌شود، می‌باشد. این رویکرد کاربردهای گسترده‌ای در حوزه‌هایی مانند بینایی ماشین، پزشکی، هنر دیجیتال و بازسازی تصاویر دارد و با بهره‌گیری از مدل‌هایی مانند شبکه‌های مولد تخاصمی (GAN) و Autoencoder و به طور کلی مدل های مولد امکان ایجاد نگاشت‌های پیچیده و معنادار بین فضاهای تصویری مختلف را فراهم می‌کند.



شکل ۱. نمونه ای از Image Translation

معماری های مهم

در این بخش با معماری‌های کاربردی که در بازسازی و تولید تصاویر وجود دارد آشنا می‌شویم.

Autoencoder (AE)

مفاهیم پایه این معماری برای اولین بار در دهه ۸۰ میلادی مطرح شد. هدف آن در ابتدا کاهش ابعاد ورودی به شکل غیرخطی بود. در زمان حال به یک مدل ریاضی پیچیده است که بر روی داده‌های بدون برچسب و بدون طبقه‌بندی آموزش می‌بیند و برای نگاشت داده‌های ورودی به یک نمایش فشرده‌شده از ویژگی‌ها استفاده می‌شود، و سپس داده‌ی ورودی را از روی آن نمایش فشرده بازسازی می‌کند، Autoencoder میگویند. این معماری دارای سه بخش اصلی است.

I. رمزگذار (Encoder)

ماژولی که داده‌های ورودی را به یک نمایش رمزگذاری‌شده فشرده تبدیل می‌کند؛ این نمایش معمولاً چندین مرتبه کوچک‌تر از داده‌های ورودی است.

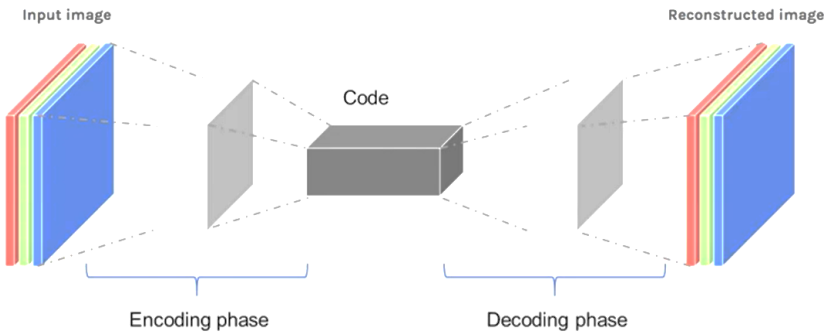
II. Bottleneck

این بخش به‌عنوان نگهبان دانش عمل می‌کند و جریان اطلاعات از رمزگذار به رمزگشا را کنترل می‌نماید. طراحی Bottleneck به‌گونه‌ای است که بیشترین اطلاعات ممکن از تصویر را ثبت می‌کند، به‌طوری که به جای حفظ کامل تصویر، یک نمایش فشرده و معنی‌دار از آن ساخته می‌شود.

III. رمزگشا (Decoder)

این فرآیند در واقع دوباره به تصویر جان می‌بخشد و در سمت دیگر شبکه، مانند یک "بازکننده‌ی فشرده‌سازی" عمل می‌کند. این بخش از یک دنباله‌ی عملیات Up-Sampling و بلوک‌های کانولوشنی برای بازسازی خروجی استفاده می‌کند.

شکل ۲، نشان دهنده معماری کلی AE است.



شکل۲. نمایش کلی Autoencoder

انواع مختلفی از این معماری وجود دارد که در کاربردهای خاصی نظیر کاهش ابعاد، نویززدایی تصویر، تشخیص ناهنجاری و ... استفاده می‌شوند. یکی دیگر از کاربردهای آن، تولید تصویر است. معماری مورد استفاده برای اینکار، Variational Autoencoders است که با نام اختصاری VAE معروف است. در ادامه به بررسی آن می‌پردازیم.

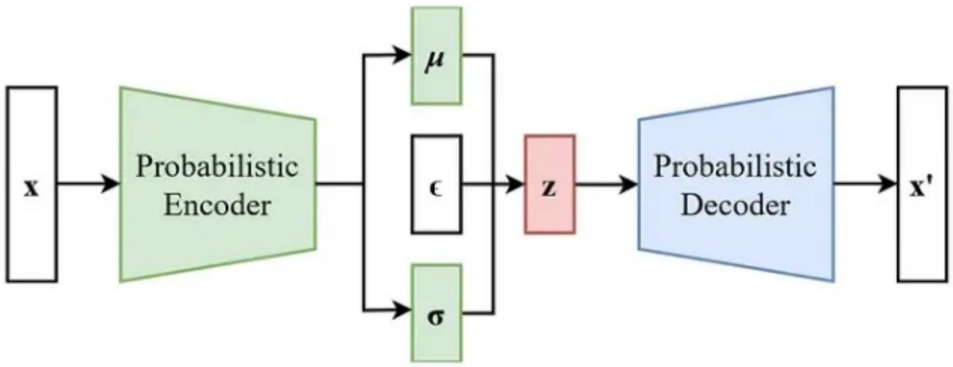
Variational Autoencoders (VAE)

یک نوع معماری شبکه عصبی مولد (Generative) است. در ساختار کلی، VAEهمان اجزای اصلی یک AE معمولی را دارد. تفاوت آن در Bottleneck آن است. Bottleneck در AE یک بردار عددی ثابت است که نمایش فشرده‌شده‌ی داده ورودی را نشان می‌دهد. اما درVAE ، نه یک بردار عددی ساده، بلکه یک توزیع آماری (معمولاً نرمال گاوسی) است. گلوگاه در VAE نه فقط یک فشرده‌ساز، بلکه یک تنظیم‌کننده‌ی آماری است که شبکه را مجبور می‌کند ویژگی‌های کلی و مهم داده را به صورت توزیعی یاد بگیرد تا هم بازسازی خوبی داشته باشد و هم بتواند داده جدید بسازد. از این رو به Bottleneck آن، Latent Space میگویند. پس خروجی Encoder یک توزیع آماری (معمولاً توزیع نرمال یا گاوسی) در فضای نهفته است و Decoder از این توزیع آماری یک نقطه تصادفی نمونه‌گیری می‌کند و تلاش می‌کند که ورودی اولیه را به‌دقت بازسازی کند.

پس مدل باید پارامترهای توزیع را یاد بگیرد. از طرفی میدانیم که فرایند یادگیری پارامترها دارای backpropagation است. علت نمونه گیری تصادفی decoder نیز به همین علت است زیرا اگر به طور مستقیم نمونه برداری کند، فرایند مشتق گیری در آن غیرقابل انجام است و به همین خاطر با یک نویز نرمال نمونه برداری انجام می‌شود. اینکار باعث میشود تابع z، نسبت به پارامترهای توزیع، دترمنیستیک باشد و فرایند تصادفی بر عهده ϵ باشد.

$$z = \epsilon . \sigma + \mu$$

شکل ۲، معماری کلی برای توزیع استاندارد را نشان میدهد.



شکل۳. معماری VAE برای توزیع گاوسین

یک معماری تولید کننده دیگر نیز وجود دارد که رویکرد متفاوتی با دو معماری قبلی دارد.

Generative Adversarial Network (GAN)

معماری GAN از دو شبکه‌ی عصبی تشکیل شده است Generator وDiscriminator، که به‌صورت هم‌زمان و از طریق یک فرایند رقابتی (Adversarial Learning) آموزش می‌بینند.

I. مولد (Generator)

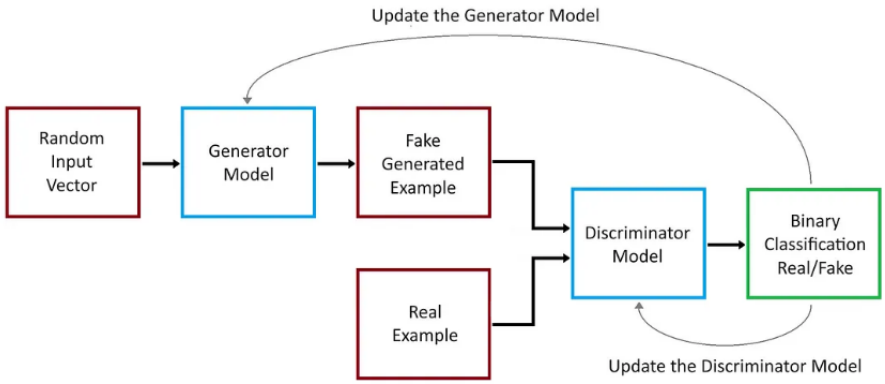
این شبکه نویز تصادفی را به‌عنوان ورودی دریافت کرده و داده‌ای مانند تصویر تولید می‌کند. هدف آن تولید داده‌ای است که تا حد امکان شبیه به داده‌های واقعی باشد.

II. متمایزکننده (Discriminator)

این شبکه داده‌های واقعی و داده‌های تولیدشده توسط مولد را به‌عنوان ورودی دریافت می‌کند و سعی دارد بین آن‌ها تمایز قائل شود. خروجی آن احتمال واقعی بودن داده‌ی ورودی است.

این دو شبکه در یک بازی رقابتی قرار دارند. مولد در تلاش است که داده‌هایی تولید کند که تمییزدهنده نتواند آن‌ها را از داده‌های واقعی تشخیص دهد، در حالی‌که تمییزدهنده سعی دارد در تمایزگذاری بین داده‌های واقعی و ساختگی بهتر عمل کند. پس اگر مولد بتواند داده های ساختگی قابل قبولی که نزدیک به داده های اصلی است تولید کند، میتوانیم بگوییم داده تولید شده خوب است.

برای اموزش این معماری، generator از یک بردار نویز تصادفی شروع به ساخت تصاویر میکند اما از توزیع داده ها اطلاعاتی ندارد. سپس در هر مرحله پس از ساخت، تصویر را به discriminator میدهد که تشخیص دهد با چه احتمالی تصویر واقعی است. از نتیجه این مولد به مرور توزیع دیتا را یاد میگیرد و این رقابت که داده تولیدی بتواند discriminator را گول بزند تبدیل به یک بازی minmax میشود.



شکل۴. معماری GAN

از کاربرد های ان میتوان به تولید تصاویر رزولوشن بالا، انتقال تصویر از یک حوزه به حوزه دیگر ، augmentation و همچنین پیش پردازش برای مدل های طبقه بند و ... است.

ارزیابی

ارزیابی مدل یکی از مهم ترین کارهایی است که عملکرد مدل را بررسی می‌کند. در تولید داده، دو نوع ارزیابی وجود دارد.

الف- کیفی:

این ارزیابی توسط متخصصان حوزه ای که مدل در آن اموزش دیده است، انجام می‌شود. مثلا اگر قصد تولید تصاویر پزشکی رزولوشن بالا را داشتیم، از چند متخصص برای ارزیابی خروجی دعوت می‌کنیم و نظرات ان‌ها را برای خروجی میپرسیم تا از اطلاعاتی از عملکرد مدل به دست بیاوریم.

ب- کمی:

در حوزه بینایی متریک هایی وجود دارند که ما میتوانیم از انها برای ارزیابی مدل استفاده کنیم. در ادامه به صورت خیلی کوتاه چند تا از آن‌ها را معرفی می‌کنیم.

- mse : میانگین مربع اختلاف بین پیکسل‌های تصویر اصلی و تصویر بازسازی‌شده
- mae : میانگین قدرمطلق خطاها بین پیکسل‌ها.
- PSNR : نشان می‌دهد که چقدر تصویر بازسازی‌شده شبیه به تصویر اصلی است (برحسب دسی‌بل).
- SSIM : کیفیت ساختاری تصویر را اندازه‌گیری می‌کند (روشنایی، کنتراست، ساختار).

در کاربردهای واقعی نیز از هر دو روش ارزیابی برای ارزیابی مدل ها استفاده می‌شود.

دیتاست

برای انجام این پروژه می‌توانید از این لینک دیتاست را دانلود کنید. انتخاب هریک از فایل های داخل لینک مجاز است. همچنین این دیتاست در سایت kaggle نیز موجود است. (انجام پروژه در kaggle نیز مجاز است.)

پروژه

استفاده از GAN برای این پروژه گزینه مناسبی است. در این مورد تحقیق کنید. (از نظر تصاویر تولیدی، پایداری و ...)

حال با توجه به توضیحات ارائه‌شده، یک مدل Image Translation برای تبدیل تصاویر موجود در دیتاست از یک حوزه به حوزه‌ای دیگر آموزش دهید.

خروجی های مورد انتظار:

۱. باید بتوانید ماسک تصویر را به عنوان ورودی در نظر گرفته و تصویر بازیابی شده آن را خروجی دهید.
دقت شود که حتما از متریک های PSNR یا SSIM استفاده شود برای ارزیابی استفاده شود.
نتیجه ارزیابی برای این دو باید حداقل برابر با:

$PSNR > 23$

$SSIM > 0.6$

۲. در این پروژه باید یک گزارش کامل و مفصل از نحوه پیاده‌سازی و عملکرد سیستم، کتابخانه‌های مورد استفاده و... ارائه دهید. در واقع هر بخش از پیاده‌سازی باید در گزارش به صورت دقیق توضیح داده شود.

نمره امتیازی:

۱. در صورت عدم استفاده از مدل های آماده می‌توانید تا ۵۰ درصد نمره امتیازی دریافت کنید.
۲. اگر از یک UI استفاده کنید، تا ۲۰ درصد نمره امتیازی دریافت می‌کنید. (دقت شود نمره ای بخش برای مدلی است که خروجی های مورد انتظار را داشته باشد.)

نکات نهایی

۱. ددلاین پروژه تاریخ ۲۶ دی است.
۲. این پروژه دارای ارائه می‌باشد و باید آمادگی ارائه آن را داشته باشید.
۳. کسب حداقل نصف نمره پروژه برای گذراندن درس الزامی است.
۴. پروژه را می‌توان به صورت انفرادی یا درگروه های دو نفره انجام داد. حتما در گزارش‌کار نام اعضا و شماره دانشجویی نوشته شود؛ در غیر این صورت نمره برای اسامی نوشته‌نشده تعلق نمی‌گیرد.
۵. در صورت ننوشتن گزارش کار، نمره‌ای به پروژه تعلق نمی‌گیرد.
۶. فایل کد و گزارش‌کار را در پوشه‌ای با نام زیر در سامانه آپلود کنید.

CV-FinalProject-std#1-std#2

موفق باشید