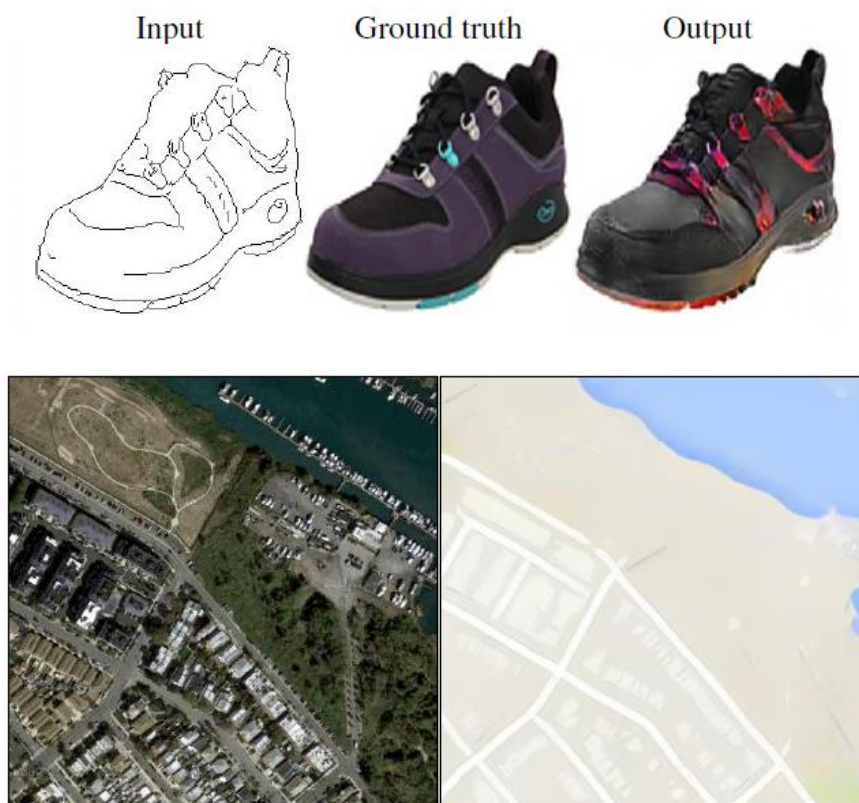




مقدمه

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 تخصصی (GAN) و Autoencoder و به طور کلی مدل‌های 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 معماری های کاربردی که در بازسازی و تولید تصاویر وجود دارد آشنا می شویم.

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 داده های بدون برچسب و بدون طبقه بندی آموزش می بیند و برای نگاشت داده های ورودی به یک نمایش فشرده شده از ویژگی ها استفاده می شود، و سپس داده ی ورودی را از روی آن نمایش فشرده بازسازی می کند، Autoencoder میگویند. این معماری دارای سه بخش اصلی است.

I. رمزگذار (Encoder)

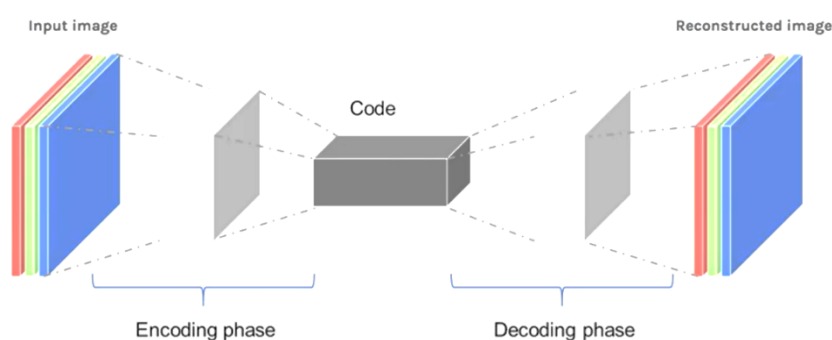
ماژولی که داده های ورودی را به یک نمایش رمزگذاری شده فشرده تبدیل می کند؛ این نمایش معمولاً چندین مرتبه کوچک تر از داده های ورودی است.

II. Bottleneck

این بخش به عنوان نگهدارنده دانش عمل می کند و جریان اطلاعات از رمزگذار به رمزگشا را کنترل می نماید. طراحی Bottleneck به گونه ای است که بیشترین اطلاعات ممکن از تصویر را ثبت می کند، به طوری که به جای حفظ کامل تصویر، یک نمایش فشرده و معنی دار از آن ساخته می شود.

III. رمزگشا (Decoder)

این فرآیند در واقع دوباره به تصویر جان می بخشد و در سمت دیگر شبکه، مانند یک "بازکننده ی فشرده سازی" عمل می کند. این بخش از یک دنباله ی عملیات Up-Sampling و بلوک های کانولوشنی برای بازسازی خروجی استفاده می کند. شکل ۲، نشان دهنده معماری کلی AE است.



شکل ۲. نمایش کلی Autoencoder

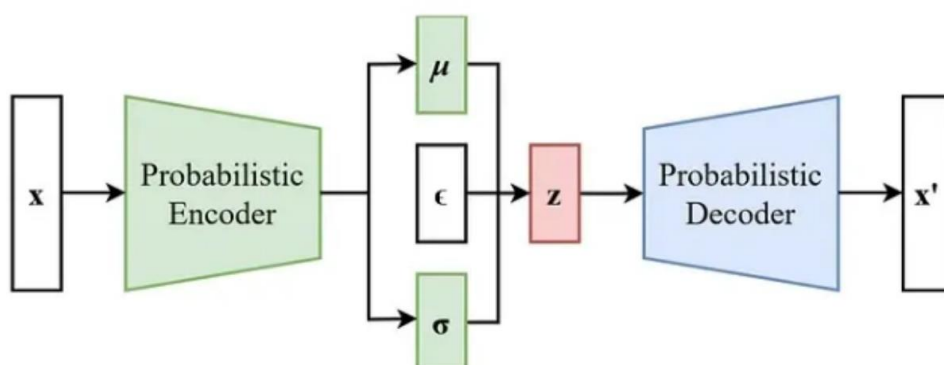
انواع مختلفی از این معماری وجود دارد که در کاربردهای خاصی نظیر کاهش ابعاد، نویززدایی تصویر، تشخیص ناهنجاری و ... استفاده می‌شوند. یکی دیگر از کاربردهای آن، تولید تصویر است. معماری مورد استفاده برای اینکار، Variational Autoencoders است که با نام اختصاری VAE معروف است. در ادامه به بررسی آن می‌پردازیم.

Variational Autoencoders (VAE)

یک نوع معماری شبکه عصبی مولد (Generative) است. در ساختار کلی، VAE همان 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 در VAE، 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 نمونه‌گیری تصادفی decoder نیز به همین علت است زیرا اگر به طور مستقیم نمونه برداری کند، فرایند مشتق‌گیری 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، دترمینیستیک باشد و فرایند تصادفی بر عهده ϵ باشد.

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

شکل ۳، معماری کلی برای توزیع استاندارد را نشان می‌دهد.



شکل ۳. 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 و Discriminator، که به صورت همزمان و از طریق یک فرایند رقابتی (Adversarial Learning) آموزش می‌بینند.

I. مولد (Generator)

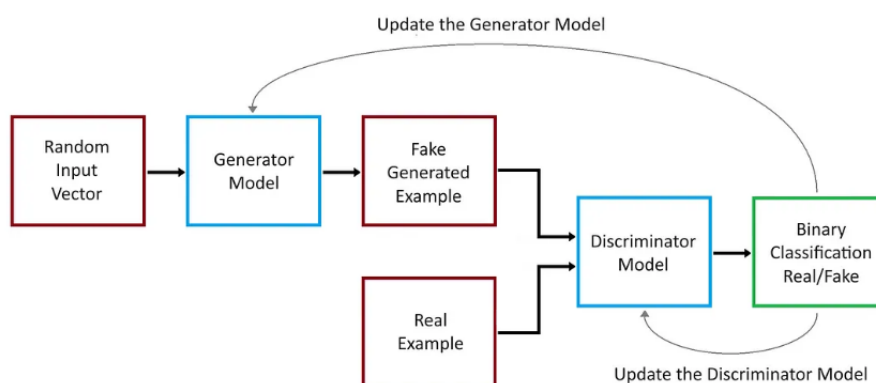
این شبکه نويز تصادفی را به عنوان ورودی دریافت کرده و داده‌ای مانند تصویر تولید می‌کند. هدف آن تولید داده‌ای است که تا حد امکان شبیه به داده‌های واقعی باشد.

II. متمایز کننده (Discriminator)

این شبکه داده‌های واقعی و داده‌های تولید شده توسط مولد را به عنوان 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 حالی که تمییزدهنده سعی دارد در تمایزگذاری بین داده‌های واقعی و ساختگی بهتر عمل کند. پس اگر مولد بتواند داده‌های ساختگی قابل قبولی که نزدیک به داده‌های اصلی است تولید کند، می‌توانیم بگوییم داده تولید شده خوب است.

برای آموزش این معماری، generator از یک بردار نويز تصادفی شروع به ساخت تصاویر میکند اما از توزیع داده‌ها 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.



شکل ۴. معماری GAN

از کاربرد های آن 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 مهم‌ترین 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 آموزش دیده است، انجام می‌شود. مثلاً اگر قصد تولید تصاویر پزشکی رزولوشن بالا را داشتیم، از چند متخصص برای ارزیابی خروجی دعوت می‌کنیم و نظرات آن‌ها را برای خروجی می‌پرسیم تا از اطلاعاتی از عملکرد مدل به دست بیاوریم.

ب - کمی:

در حوزه بینایی متریک‌هایی وجود دارند که ما می‌توانیم از آنها 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 مدل‌ها استفاده می‌شود.

دیتاست

برای انجام این پروژه می‌توانید از این [لینک](#) دیتاست را دانلود کنید. انتخاب هریک از فایل‌های 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 تصاویری که بخشی از آن مخدوش شده را ورودی 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 >

SSIM >

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 گزارش کار نام اعضا و شماره دانشجویی نوشته شود؛ در غیر این صورت نمره برای اسامی نوشته نشده تعلق نمی‌گیرد.

۵. در صورت ننوشتن گزارش کار، نمره‌ای به پروژه تعلق نمی‌گیرد.

۶. فایل کد و گزارش کار را در پوشه‌ای با نام زیر در سامانه آپلود کنید.

CV-Final-Project-std#1-std#2