

توضیح دیتاست مجموعه داده‌هایی در اختیار شما قرار دارد که شامل تصاویر حروف بزرگ الفبای انگلیسی از ۹۰۰۰ فونت مختلف است. در این دیتاست، اسم هر فایل نشان‌دهنده این است که آن تصویر مربوط به چه حرف و از چه فونتی است. به عنوان مثال فایل F_00312، تصویر مربوط به حرف F در فونت ۳۱۲م می‌باشد. بنابراین تصویری که شماره‌های یکسانی دارند از یک فونت هستند. مثلاً تصاویر فونت شماره ۲۲ به صورت زیر هستند:

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z

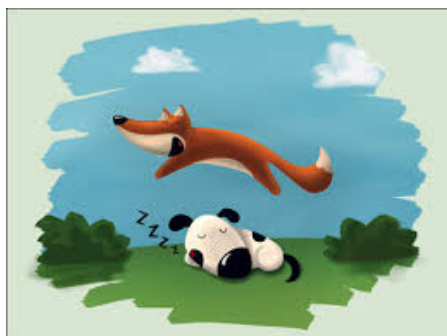
قسمت اول: طراحی دسته‌بند در این قسمت با استفاده از پایگاه داده‌ای که در اختیار دارید یک دسته بند با استفاده از شبکه‌های عصبی کانولوشنی آموزش دهید. این شبکه باید تصویر ورودی 32×32 در ورودی بگیرد و در خروجی دسته متناظر (کاراکتر متناظر) را از بین ۲۶ حرف الفبای انگلیسی مشخص کند. برای ساختار شبکه می‌توانید از هر شبکه کانولوشن به اختیار خود استفاده کنید. (۴۰ نمره)

ارزیابی قسمت اول: برای ارزیابی دسته بند شما، به شما یک ورودی به صورت $1000 \times 32 \times 32$ داده خواهد شد (۱۰۰۰ تصویر از حروف الفبا) و از شما می‌خواهیم برچسب مربوط به هر تصویر را تولید نمائید و به صورت یک بردار 1000 بعدی در خروجی تحویل دهید. برای این منظور فایل با نام test1.py آماده نمایم که در آن یک فایل با نام x_test.json را load کرده و x_test را از آن بگیرد (که یک tuple با ابعاد گفته شده است) و در نهایت یک فایل json با نام prediction.json را که شامل یک بردار 1000 بعدی است، ذخیره نماید. دقت کنید که ابعادهای بردار خروجی بایستی به صورت کاراکتر باشد که این کاراکترها یکی از "A" تا "Z" خواهند بود.

برای رسیدن به نمره کامل این قسمت باید دقت شما روی داده‌های تست ۹۰٪ باشد. افرادی که دقت بیشتر از آن به دست بیاورند به صورت رقابتی نمره امتیازی این بخش را به دست می‌آورند. (۱۵ نمره) لازم به ذکر است که منظور از دقت، نسبت داده‌های درست طبقه‌بندی شده به تعداد کل داده‌ها است.

قسمت دوم: شبکه مولد در این قسمت هدف پیاده‌سازی یک شبکه مولد شرطی (Conditional GAN) است. شبکه مولد باید برای هر حرف که در ورودی شبکه داده می‌شود یک تصویر جدید از آن حرف تولید کند. شبکه متمایزساز نیز باید با استفاده از تصویری که به آن داده می‌شود و همچنین نوع کاراکتر که در ورودی خود می‌گیرد، واقعی یا ساختگی بودن تصویر را تشخیص دهد. نویز مورد نیاز برای تابع مولد را از توزیع نرمال $N(0, I)$ بگیرید. (ابعاد آن را به عنوان یک هاپرپارامتر تنظیم کنید. توجه کنید که هرچه ابعاد نویز در شبکه پیشنهادی کمتر باشد نمره بیشتری می‌گیرید). پس از پایان یادگیری تعدادی از نمونه کاراکترهای تولیدی را ذخیره و کنید در گزارش خود بیاورید (برای هر حرف ۱۰ تصویر در گزارش خود بیاورید). همچنین در گزارش خود نمودار تابع هزینه برای هر دو شبکه مولد و متمایزساز را ارائه کنید. همچنین جمله زیر که شامل تمامی حروف الفبای انگلیسی است را با کاراکترهای تولید شده توسط شبکه مولد خود تولید کنید. (۶۰ نمره)

THE QUICK BROWN FOX JUMPS OVER A LAZY DOG



ارزیابی قسمت دوم: برای ارزیابی این قسمت قصد داریم با تولید جمله بالا، کیفیت شبکه مولد شما را بررسی می‌شود. به این منظور شما باید یک

کد با نام test2.py آماده نمایید که در آن با تولید بردار تصادفی، هر یک از حروف بالا را با کمک شبکه مولد (GAN) خود تولید نموده و حاصل را به صورت یک تصویر ۱۳۱۲×۳۲ (با احتساب فاصله ها ۴۱ حرف) که از کنار هم قرار دادن تصاویر ۳۲×۳۲ است با نام Sentence.png ذخیره نمایید. دقت کنید که در مدل های GAN ممکن است به ازای بعضی از نقاط تصادفی حاصل خروجی کیفیت لازم را نداشته باشد. برای جلوگیری از این مشکل، شما می توانید به ازای هر کاراکتر از جمله بالا ۵ یا ۱۰ بردار latent تصادفی تولید کرده و با کمک آن حروف را تولید نمایید و از میان ۵ یا ۱۰ کاراکتر به دست آمده، با کمک دسته بندی که در قسمت قبل آموزش دادید، تصمیم بگیرید که از کدام یک برای ساختن جمله استفاده نمایید. به عنوان مثال برای تولید حرف T در اول جمله، کد باید به این صورت عمل کند که ۱۰ بردار کاملاً تصادفی (بدون مقاردهی اولیه برای مولد عدد تصادفی) تولید نموده و با هر یک از آن ها یک T تولید کند. سپس آن ها را به دسته بند آموزش داده شده در قسمت قبل بدهد و آن تصویری را انتخاب کند که مقدار احتمال برای T آن بیشینه است. به کمک این روش می توانید تصاویر تولید شده توسط شبکه مولد را در حالت ایده آل آن نمایش دهید.

قسمت امتیازی ۱: در این مقاله چندین معیار برای ارزیابی داده های تولید شده توسط شبکه GAN معرفی شده است. می توانید حداکثر دو معیار را انتخاب نموده و آن را برای مدل خود بررسی نمایید و در گزارش ذکر کنید. (۵ نمره)

قسمت امتیازی ۲: شبکه مولدی (GAN) پیاده سازی کنید که در خروجی یک فونت تولید کند. بدین معنی که تمامی حروف تولید شده در خروجی شبکه مولد باید از یک فونت باشند. در گزارش خود تعدادی فونت تولید شده را رسم کنید و نمودار تابع هزینه های به کار رفته برای آموزش شبکه را نیز ارائه کنید. (۲۵ نمره)