به نام خدا

تمرین سری دوم درس رباتیک اجتماعی و شناختی

(شبکه های عصبی پیچشی)

مدرس: علیرضا طاهری

دانشکده مهندسی مکانیک، دانشگاه صنعتی شریف

۱- همانطور که اطلاع دارید، شبکه های عصبی عمیق قادر به تولید داده های مصنوعی نظیر تصویر، صدا، ویدئو و حرکت های غیر واقعی (و البته بعضاً غیر قابل تشخیص برای انسان ها) نیز می باشند. از سایت https://thispersondoesnotexist.com با ریفرش های متوالی، تعدادی عکس (حداقل ۳۰ تصویر) ذخیره کرده و میانگین آن ها را رسم کنید. آیا تصویر به دست آمده از میانگین گیری شما، چهره فرد خاصی است؟

۲- با جستجوی اینترنتی و به صورت کیفی به پرسش های زیر پاسخ دهید:

الف) هنگامی که در فرآیند آموزش شبکه های عصبی از drop-out بهره می گیریم، دیگر در زمان تست از drop-out استفاده نکرده و در عوض وزن یال های خروجی را در p (keep probability) ضرب می کنیم. علت این امر را توضیح دهید.

ب) ممکن است از اسلایدهای درس به خاطر داشته باشید که در شبکه های عصبی عیمق (نظیر ResNet) بعضاً از ماژولی به نام (Batch-Normalization (BN استفاده شده بود. تحقیق کنید BN چگونه به آموزش شبکه های عصبی کمک می کند؟

۳- شبکه عصبی پیچشی زیر را در نظر بگیرید.

Conv۲D(۶۴(۵*۵)، stride=(۱،۱))، Relu
Conv۲D(۶۴(۵*۵)، stride=(۱،۱))، Relu، max pooling((۲*۲)، stride=(۲،۲))
Conv۲D(۶۴(۳*۳)، stride=(۱،۱))، Relu، max pooling((۲*۲)، stride=(۲،۲))
Fully connected(number of outputs=۲۵۶)، Relu
Fully connected(number of outputs=number of class)، Softmax

الف) با فرض ورودی ۳×۳۲×۳۲ و تعداد کلاس ۱۰ تعداد پارامترهای شبکه را تعیین کنید.

ب) شبکه داده شده را برای داده های Cifar10 پیاده سازی کرده و نمودار دقت و تابع هزینه را گزارش دهید (این دیتاست در کراس موجود است).

پ) با استفاده از مفهوم Transfer Learning می خواهیم وزن های یادگرفته شده در شبکه بخش ب را برای آموزش روی ۴ کلاس از پایگاه داده Cifar100 (به عنوان نمونه: bicycle, mountain, table, dolphin) استفاده کنیم. برای این منظور از وزن های یادگرفته شده لایه های پیچشی بخش ب استفاده کنید که فریز شده و آپدیت نمی شوند؛ ولی لایه های

fully connected آن قابل آموزش و تغییر می باشند. طبیعتاً در مسئله ذکر شده، لایه آخر باید ۴ نرون معادل ۴ کلاس ذکر شده، داشته باشد. نمودار دقت و تابع هزینه را گزارش دهید (این دیتاست نسخه کامل تر Cifar10 و دارای ۱۰۰ طبقه بوده و در کراس موجود است).

ت) یک شبکه غیر ترتیبی (منظور با بهره گیری از Functional API در کراس) با معماری دلخواه (خواه کاملاً دست ساز و یا با بهره گیری ترکیبی از ماژول/ایده های اینسپشن، ResNet و ...) برای طبقه بندی دادگان Cifar10 ارائه دهید. تصویر معماری شبکه های خود را رسم کرده و ماتریس درهم ریختگی را بر روی دادگان تست نمایش دهید.

۴- اجازه دهید مجدداً به سوال پایانی تمرین سری قبل در زمینه تشخیص حالت چهره افراد (نظیر شادی، غم، خشم و ...) بر اساس مجموعه عکس هایی مستخرج از پایگاه داده CK+ بازگردیم.

الف) شما می بایست یک شبکه عصبی پیچشی طراحی کنید که تا حد امکان، طبقه بندی ۸ کلاسه شناسایی حالت های چهره کاربران را به خوبی انجام دهد. عکس ها در هشت پوشه مختلف قرار داده شده اند و نام پوشه ها، معادل برچسب عکس موجود در آن ها است. ما در نهایت کدهای شما را روی مجموعه دادگانی به عنوان داده های تست (که هم اکنون در اختیار شما نیست) نیز پیاده سازی خواهیم کرد تا میزان توانمندی شبکه پیشنهادی شما را در طبقه بندی داده های جدید محک بزنیم.

لینک دانلود داده های آموزش:

https://drive.google.com/file/d/1XL0W5YdJjdE4WVfo9YHjddH4V6CJ6H10/view?usp=sharing

ب) یک عکس از این پایگاه داده انتخاب کنید. به دلخواه، یک یا تعدادی از فیلترهای لایه های پیچشی و همچنین نتیجه اعمال آن فیلترها بر عکس انتخابی خود را به صورت گرافیکی رسم نمایید. آیا فیلتر های انتخابی شما در حال آموزش ویژگی خاص و قابل بیانی از تصویر می باشند؟

پ) شبکه نهایی خود را جهت ارزیابی میزان دقت شناسایی حالت چهره (خود یا اطرافیان) از طریق وب کم لپ تاپ آزمایش کرده و نتیجه را گزارش دهید (بد نیست فیلمی در حد یک دقیقه از عملکرد این شبکه بر روی چهره خودتان را نیز بارگذاری کنید). ورودی شبکه تان عکس های گرفته شده در هر فریم توسط وب کم است.

علاوه بر خواسته های مسائل، می بایست تمامی کدهای خود را نیز بارگذاری کنید.