

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

گزارش کار تمرین ۴-۳ - یادگیری عمیق

عنوان تمرین :

طراحی و پیاده‌سازی شبکه عصبی تمام متصل (FCN)

برای طبقه‌بندی تصاویر CIFAR-10

دانشجو : محمدداود وهاب رجائی

شماره دانشجویی : ۴۰۴۱۴۱۹۰۴۱

استاد محترم : دکتر کیوان راد

حل تمرین : دکتر گلیزاده

فهرست مطالب

چکیده	۱
مقدمه و بیان مسئله.....	۱
روش‌شناسی (Methodology)	۲
۱. آماده‌سازی داده‌ها.....	۲
۲. معماری مدل پیشنهادی	۳
۳. پیکربندی آموزش	۴
نتایج تجربی(Experimental Results)	۴
۱. روند آموزش و اعتبارسنجی	۴
۲. ارزیابی روی داده‌های تست.....	۵
۳. تحلیل جزئیات کلاس‌ها.....	۵
بحث و تحلیل (Discussion).....	۶
۱. تحلیل ماتریس در هم ریختگی.....	۶
۲. محدودیت‌های معماری FCN	۶
نتیجه‌گیری و کارهای آتی	۷

چکیده

در این تمرین، یک شبکه عصبی تمام متصل (Fully Connected Neural Network) برای طبقه‌بندی تصاویر رنگی مجموعه داده CIFAR-10 طراحی و آموزش داده شد. هدف اصلی، بررسی کارایی معماری‌های MLP بر روی داده‌های تصویری و تأثیر تکنیک‌های تنظیم و افزایش داده (Data Augmentation) بود. مدل نهایی با معماری ۳ لایه و استفاده از Dropout و تابع فعال‌ساز LeakyReLU پیاده‌سازی شد. همچنین برای بهبود تعمیم‌پذیری، از تکنیک‌های تغییرات هندسی روی داده‌های آموزشی استفاده گردید. پس از ۴۰ دور آموزش (Epoch)، مدل به دقت نهایی ۵۶.۷۶٪ روی داده‌های تست یافت. نتایج نشان داد که استفاده از Augmentation علی‌رغم سخت‌تر کردن فرآیند آموزش، منجر به یادگیری ویژگی‌های پایدارتر و عملکردی بسیار بالاتر از حدس تصادفی (۱۰٪) شده است.

مقدمه و بیان مسئله

طبقه‌بندی تصاویر یکی از مسائل بنیادین در بینایی ماشین است. در حالی که شبکه‌های کانولوشنی (CNN) استاندارد طلایی برای این وظیفه محسوب می‌شوند، درک عملکرد و محدودیت‌های شبکه‌های تمام متصل (FCN) به عنوان یک پایه (Baseline) ضروری است.

در این تمرین، چالش اصلی تبدیل داده‌های تصویری دو بعدی به بردارهای یک بعدی و آموزش یک شبکه عمیق برای یادگیری الگوهای بصری بدون استفاده از فیلترهای کانولوشنی است. ما از دیتابست CIFAR-10 شامل ۱۰ کلاس (هوایپیما، خودرو، پرنده، گربه، گوزن، سگ، قورباغه، اسب، کشتی و کامیون) استفاده کردیم که شامل ۶۰,۰۰۰ تصویر رنگی 32×32 پیکسل است. علاوه بر معماری پایه، در این پژوهش از تکنیک‌های پیشرفته‌تری مانند افزایش داده (Data Augmentation) برای غلبه بر محدودیت‌های ذاتی شبکه‌های تمام متصل و بهبود دقت نهایی بهره گرفته شده است.

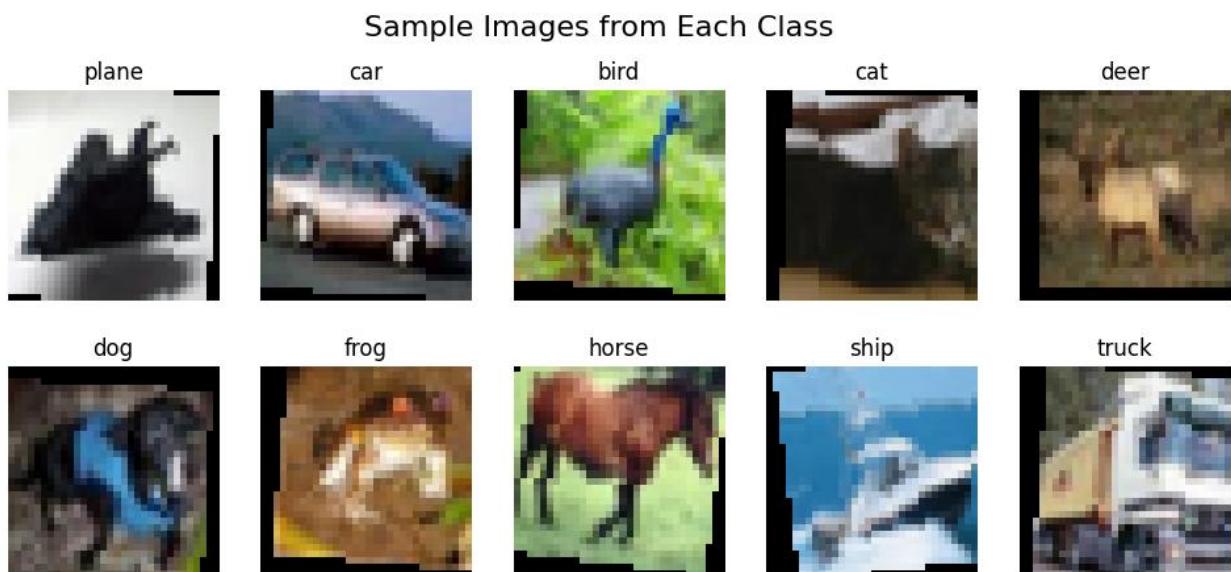
روش شناسی (Methodology)

۱. آماده سازی داده ها

- پیش پردازش (Preprocessing) : تمامی تصاویر ابتدا به تانسورهای PyTorch تبدیل شدند. جهت کمک به همگرایی سریع تر مدل و پایداری گرادیان ها، نرمال سازی (Normalization) با میانگین ۰.۵ و انحراف معیار ۰.۵ روی کanal های رنگی اعمال شد تا مقادیر پیکسل ها در بازه [۰,۱] قرار گیرند.
- تقسیم بندی داده ها :

 - داده های آموزش (Training Set) (۴۲,۵۰۰ تصویر٪ ۸۵) :
 - داده های اعتبارسنجی (Validation Set) (۷,۵۰۰ تصویر٪ ۱۵) جهت تنظیم های پردازشی.
 - داده های تست (Test Set) (۱۰,۰۰۰ تصویر جهت ارزیابی نهایی).

- بارگذاری : از DataLoader با استفاده از Batch Size (۱۲۸) برابر با استفاده شد.



شکل ۱ : نمونه تصاویر مجموعه داده CIFAR-10 پس از نرمال سازی

- افزایش داده (Data Augmentation) : به منظور بهبود تعمیم‌پذیری مدل و مقابله با بیش‌برازش، از تکنیک‌های افزایش داده روی مجموعه آموزشی استفاده شد. این تکنیک‌ها شامل چرخش‌های جزئی و وارونه‌سازی افقی (Horizontal Flip) تصاویر هستند تا مدل نسبت به تغییرات هندسی مقاوم‌تر شود.

۲. معماری مدل پیشنهادی

مدل طراحی شده یک شبکه چندلایه پرسپترون (MLP) با مشخصات زیر است :

۱. لایه ورودی : تصویر $32 \times 32 \times 3$ باز شده (Flatten) و به یک بردار با ابعاد 3072×1 تبدیل می‌شود.

۲. لایه مخفی اول : 512 نورون + LeakyReLU + Batch Normalization + Dropout (0.3)

۳. لایه مخفی دوم : 256 نورون + Batch Normalization + LeakyReLU + Dropout (0.3)

۴. لایه مخفی سوم : 128 نورون + LeakyReLU + Batch Normalization + Dropout (0.3)

۵. لایه خروجی : لایه خطی با 10 نورون (متناظر با 10 کلاس).

- تعداد کل پارامترهای قابل آموزش: $1,708,810$ پارامتر.
- دلیل انتخاب : استفاده از LeakyReLU برای جلوگیری از مشکل "مرگ نورون‌ها" و استفاده از Dropout سنگین (0.5) برای مقابله با بیش‌برازش شدید در شبکه‌های تمام متصل.

۳. پیکربندی آموزش

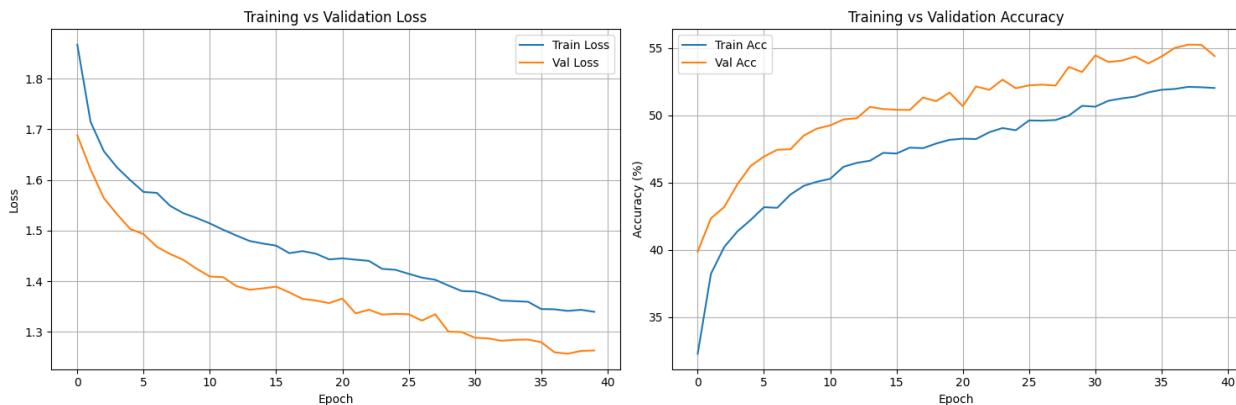
- تابع هزینه : CrossEntropyLoss
- بهینه‌ساز Adam: با نرخ یادگیری اولیه ۰.۰۰۱
- زمان‌بندی (Scheduler): استفاده از ReduceLROnPlateau برای کاهش نرخ یادگیری در صورت عدم بهبود دقت اعتبارسنجی.
- تعداد اپک : ۴۰ دور کامل.

نتایج تجربی (Experimental Results)

۱. روند آموزش و اعتبارسنجی

فرآیند آموزش در طول ۴۰ اپک نشان‌دهنده یک همگرایی پایدار بود. استفاده از Data Augmentation باعث شد در اپوک‌های ابتدایی، نرخ خطأ با نوسان بیشتری کاهش یابد (که طبیعی است)، اما در نهایت منجر به یادگیری ویژگی‌های عمیق‌تر شد. همانطور که در نمودار زیر مشهود است، فاصله میان دقت آموزش و اعتبارسنجی بسیار کم است که نشان‌دهنده کنترل عالی پدیده Overfitting و Dropout توسط Augmentation می‌باشد.

مدل نهایی به بهترین دقت اعتبارسنجی ۵۵.۲۴٪ دست یافت و ذخیره شد.



شکل ۲: منحنی‌های یادگیری؛ روند تغییرات Loss و Accuracy در طول ۴۰ اپک

۲. ارزیابی روی داده‌های تست

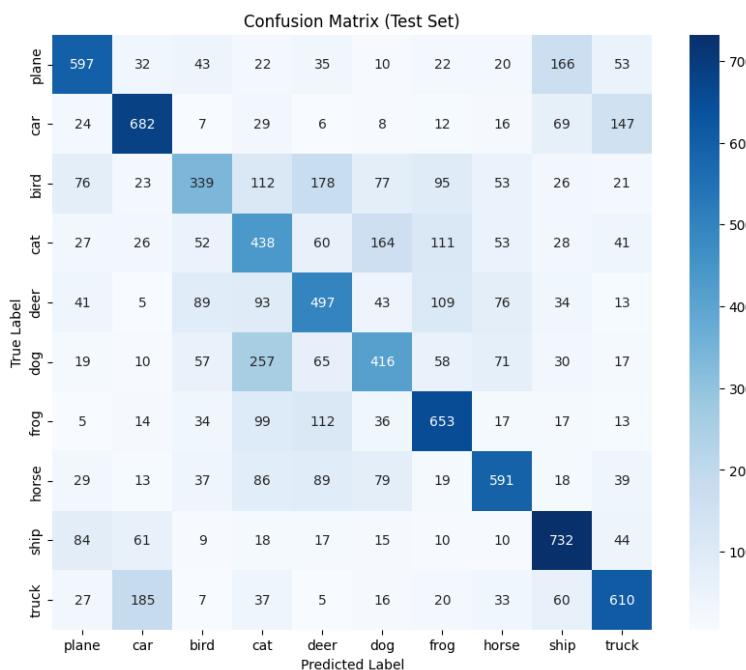
پس از پایان آموزش، مدل نهایی روی ۱۰,۰۰۰ تصویر دیده نشده تست ارزیابی شد :

- دقตنهایی (Test Accuracy) : ۵۶.۷۶%
- نکته قابل توجه این است که دقت روی داده‌های تست (۵۶.۷۶%) از دقت اعتبارسنجی (۵۵.۲۴%) بالاتر است. این پدیده نشان‌دهنده اثر مثبت Data Augmentation در جلوگیری از حفظ کردن داده‌ها و یادگیری ویژگی‌های عمومی و قابل تعمیم است.

۳. تحلیل جزئیات کلاس‌ها

با بررسی گزارش طبقه‌بندی (Classification Report)، نتایج زیر حاصل شد :

- بهترین عملکرد: کلاس‌های "کشتی" (Ship) با F1-Score برابر ۰.۶۷ و "خودرو" (Car) با ۰.۶۷.
- ضعیفترین عملکرد: کلاس‌های "گربه" (Cat) با F1-Score برابر ۰.۴۰ و "پرنده" (Bird) با ۰.۴۱.



شکل ۳ : ماتریس درهم‌ریختگی نتایج تست

بحث و تحلیل (Discussion)

۱. تحلیل ماتریس در هم ریختگی

با دقت در ماتریس در هم ریختگی (شکل ۳) مشاهده می شود که بیشترین اشتباها مدل بین کلاس های زیر رخ داده است:

۱. گربه و سگ: مدل اغلب این دو حیوان را با هم اشتباه می گیرد. دلیل این امر شباهت ساختاری (چهارپا بودن، داشتن گوش و دم) و عدم توانایی شبکه FCN در استخراج ویژگی های ظریف محلی است.

۲. پرنده و هوایپیما: اشتباہ بین این دو کلاس اغلب به دلیل رنگ پس زمینه (آسمان آبی) رخ می دهد. از آنجا که شبکه تمام متصل به شدت به مقادیر پیکسل ها و رنگ وابسته است، بافت کلی تصویر را ملاک قرار می دهد.

۲. محدودیت های معماری FCN

دستیابی به دقت ۵۵٪ برای یک شبکه تمام متصل روی دیتاست پیچیده ای مثل CIFAR-10 نتیجه ای قابل قبول است.

دلیل عدم دستیابی به دقت های بالاتر (مانند ۹۰٪ در CNN ها)، از بین رفتن اطلاعات مکانی (Spatial Information) هنگام مسطح کردن (Flatten) تصویر است. شبکه FCN رابطه همسایگی بین پیکسل ها را درک نمی کند و نمی تواند ویژگی های مستقل از مکان (Translation Invariance) را یاد بگیرد.

نتیجه‌گیری و کارهای آتی

در این تمرین، با طراحی یک شبکه عصبی عمیق ۳ لایه و بهره‌گیری از تکنیک‌های Dropout، Batch Normalization و Data Augmentation موفق به کسب دقیق ۵۶.۷۶٪ روی دیتاست CIFAR-۱۰ شدیم. نتایج نشان داد که افزایش داده (Augmentation) نقش کلیدی در بهبود تعمیم‌پذیری مدل ایفا کرده و معما ری پیشنهادی علی‌رغم محدودیت‌های ذاتی شبکه‌های تمام‌متصل، عملکرد قابل قبولی ارائه داده است.

پیشنهادات برای کارهای آتی :

استفاده از شبکه‌های کانولوشنی (CNN) برای حفظ اطلاعات مکانی و استخراج ویژگی‌های بصری عمیق‌تر.

استفاده از Data Augmentation (مانند چرخش، برش و تغییر رنگ تصادفی) برای افزایش تنوع داده‌های آموزشی و بهبود تعمیم‌پذیری مدل.

با تشکر از نگاه پر مهر شما