گزارش تیمی مدل‌های طبقه‌بندی تصویر توپ‌های ورزشی

(مقدسین-نوری-گودرزی)

مدل‌های استفاده‌شده

مدل‌های استفاده‌شده در این پروژه شامل EfficientNetB0 ، EfficientNetB2، ResNet18 ,ResNet50 در دو حالت با fine-tuning و بدون fine-tuning و همچنین گاهی با hyper parameter tuning جزئی، و همچنین VGG16بودند. همه این مدل‌ها با روش transfer learning آموزش داده شدند تا بتوانند از دانش قبلی مدل‌های از پیش‌آموزش‌دیده استفاده کنند و عملکرد بهتری در طبقه‌بندی تصاویر توپ‌های ورزشی داشته باشند.

پیش پردازش و تنظیمات مدل

در مدل‌های فوق معمولا از Image augmentation , resize , Normalization برای پیش پردازش ا ستفاده شده است. همچنین در Image augmentation روش‌های چرخش، زوم و چرخش تصادفی افقی به کار گرفته شده اند. در تنظمات مدل ها غالبا سعی شده از بیش برازش مدل با استفاده از تنظیم برخی پارامترها جلوگیری شود. همچنین تعداد ایپاک در برخی اجراها با توجه به منابع در دسترس تنظیم شده است. لازم بذکر است در برخی مدل ها سعی شده است جهت جلوگیری از میرایی گرادیان، نرخ یادگیری به صورت پویا تنظیم شود.

نتایج کلی

در بخش نتایج کلی، عملکرد مدل‌ها روی تست داخلی نشان داد که ResNet50 fine-tuned بهترین نتیجه را با دقت حدود ۰٫۹۲ به دســـت آورد. مدل‌های ResNET18 , EfficientNetB0 و EfficientNetB2 نیز عملکرد قابل قبولی داشتند و دقت آن‌ها در بازه حدود ۰٫۸۲ تا ۰٫۸۸ قرار گرفت. در مقابل، VGG16 ضعیف‌ترین عملکرد را در میان مدل‌های بررسی شده داشت و دقت آن حدود ۰٫۷۵ بود. در تست جدید نیز کاهش دقت در تمام مدل‌ها مشاهده شد و برخی کلاس‌ها، به‌ویژه در تصاویر پیچیده‌تر، افت عملکرد بیشتری نشان دادند.

طبق جدول ۱، مهم‌ترین نتیجه پروژه اختلاف بین عملکرد روی دو دیتاست است:

جدول ۱- نتایج مدل‌ها بر روی دو مجموعه داده تست

ردیف	مدل	میزان دقت روی مجموعه تست اصلی (٪)	میزان دقت روی مجموعه تست جدید (٪)	میزان افت دقت در تست جدید (٪)
۱	EfficientNetB0 با Fine-tuning چهار لایه	۸۷٫۱۳	۷۵٫۶۳	۱۱٫۵۰
۲	EfficientNetB0 با Fine-tuning پنج لایه	۸۸٫۵۹	۷۵٫۶۳	۱۲٫۹۶
۳	EfficientNetB0 بدون Fine-tuning	۷۷٫۴۶	۶۳٫۰۳	۱۴٫۴۳
۴	EfficientNetB2 بدون Fine-tuning	۸۲٫۳۰	۶۸٫۹۱	۱۳٫۳۹

گزارش تیمی مدل‌های طبقه‌بندی تصویر توپ‌های ورزشی

(مقدسین-نوری-گودرزی)

۵	ResNet50 بدون Fine-tuning	۸۷,۰۰	۸۰,۶۷	۶,۳۳
۶	ResNet50 با Fine-tuning	۹۲,۰۰	۷۷,۳۱	۱۴,۶۹
۷	ResNet18 بدون Fine-tuning	۸۵	۷۱,۴۳	۱۳,۵۷
۸	VGG16 بدون Fine-tuning	۷۵,۹۴	۵۴,۲۴	۲۱,۷

جدول ۲- جمع‌بندی نتایج

معیار مقایسه	بهترین مدل
بیشترین دقت روی مجموعه تست اصلی	ResNet50 با Fine-tuning
بیشترین دقت روی مجموعه تست جدید	ResNet50 بدون Fine-tuning
کمترین افت دقت بین دو مجموعه تست	ResNet50 بدون Fine-tuning
ضعیف‌ترین عملکرد روی مجموعه جدید	VGG16 بدون Fine-tuning

تحلیل رفتار کلاس ها

کلاس‌های ضعیف (بیشترین خطا):

۱. **Hockey Ball** ضعیف‌ترین کلاس در تقریباً همه مدل‌ها (اغلب با Golf Ball و Hockey Puck اشتباه گرفته می‌شود)
۲. **Hockey Puck** با Rugby Ball و Billiard Ball اشتباه دارد.
۳. **Shuttlecock** با Table Tennis Ball و Tennis Ball اشتباه می‌شود.
۴. **Rugby Ball** و **American Football** گاهی با هم یا با Football اشتباه می‌شوند.
۵. **Table Tennis Ball** با Tennis Ball و Golf Ball همپوشانی دارد.

علت اصلی: شباهت ظاهری زیاد بین توپ‌ها (اندازه، رنگ، شکل و بافت) به‌خصوص در مدل‌هایی که Fine-Tuning ضعیف یا متوسط داشتند.

کلاس‌های قوی (دقت بالا):

- **Basketball**
- **Football (Soccer)**
- **Volleyball**

- **Tennis Ball**
- **Billiard Ball**
- **Golf Ball**

تحلیل خطا

خطاهای اصلی در مدل‌ها شامل موارد ذیل می‌باشند:

- ☒ اشتباه بین کلاس‌های مشابه
- ☒ حساسیت مدل‌ها به پس‌زمینه‌های شلوغ
- ☒ افت عملکرد در تصاویر کم‌کیفیت یا زاویه‌دار
- ☒ ناترازی نسبی تعداد داده ها در کلاس های مختلف که باعث سوگیری مدل آموزش دیده می شود.

در داده تست داخلی، عملکرد خوبی مشاهده می شود؛ اما به هنگام استفاده از داده تست خارجی (جدید) افت عملکرد پدیدار می‌گردد. این موضوع، نشان‌دهنده وجود domain shift و وابستگی مدل به دیتاست آموزشی است.

پیشنهاد برای بهبود خروجی مدل

جهت بهبود دقت مدل‌های استفاده شده می‌توان از راه‌کارهای ارائه‌شده ذیل استفاده نمود:

- ✓ افزایش داده‌های واقعی (**مهم‌ترین عامل**)
- ✓ تقویت data augmentation
- ✓ استفاده از fine-tuning عمیق‌تر و hyper parameter tuning بیشتر برای ResNet
- ✓ استفاده از ensemble models
- ✓ بررسی confusion matrix برای کلاس‌هایی که مدل در تشخیص آن، بیش‌تر دچار مشکل می‌شود
- ✓ بالانس کردن تعداد داده‌های آموزش در کلاس های مختلف

جمع بندی

مدل‌های آموزش‌دیده برای یک prototype مناسب هستند، اما برای استفاده صنعتی هنوز نیاز به بهبود دارند. مشکل اصلی معماری مدل نیست، بلکه تفاوت بین داده آموزشی و شرایط واقعی است.