

قطعه‌بندی تصاویر حیوانات خانگی با استفاده از شبکه عصبی عمیق U-Net

۱. مقدمه و بیان مسئله

قطعه‌بندی معنایی تصاویر (Semantic Segmentation) یکی از مسائل مهم در حوزه بینایی ماشین است که هدف آن، تخصیص یک برچسب معنایی به هر پیکسل تصویر می‌باشد. برخلاف مسائل دسته‌بندی که تنها یک برچسب کلی به تصویر اختصاص داده می‌شود، در قطعه‌بندی، ساختار مکانی اشیاء نیز باید به‌دقت حفظ شود.

در این پروژه، مسئله‌ی قطعه‌بندی تصاویر حیوانات خانگی (سگ و گربه) مورد بررسی قرار گرفته است. تشخیص دقیق ناحیه‌ی حیوان از پس‌زمینه می‌تواند در کاربردهایی مانند تحلیل تصاویر حیوانات، پردازش داده‌های بصری و توسعه سامانه‌های بینایی ماشین مورد استفاده قرار گیرد.

هدف این پروژه، طراحی، پیاده‌سازی و ارزیابی یک مدل یادگیری عمیق مبتنی بر معماری U-Net برای قطعه‌بندی پیکسلی تصاویر است.

۲. معرفی دیتاست

در این پروژه از دیتاست Oxford-IIIT Pet Dataset استفاده شده است. این دیتاست شامل تصاویر رنگی از سگ‌ها و گربه‌ها به همراه ماسک‌های قطعه‌بندی پیکسلی می‌باشد.

ماسک هر تصویر شامل سه کلاس است:

- Background (پس‌زمینه)
- Pet (بدنه حیوان)
- Border (مرز اطراف حیوان)

داده‌ها به نسبت ۸۰٪ برای آموزش و ۲۰٪ برای اعتبارسنجی تقسیم شده‌اند. در شکل ۱ نمونه‌هایی از تصاویر ورودی و ماسک‌های قطعه‌بندی متناظر نمایش داده شده است.



شکل ۱ - نمونه‌هایی از تصاویر ورودی و ماسک‌های قطعه‌بندی دیتاست Oxford-IIIT Pet

۳. پیش‌پردازش داده‌ها

به‌منظور آماده‌سازی داده‌ها برای آموزش شبکه عصبی، مراحل پیش‌پردازش زیر انجام شده است:

- تغییر اندازه تصاویر به 256×256
- اعمال افزایش داده (Data Augmentation) شامل:
 - چرخش افقی تصادفی
 - تغییر تصادفی روشنایی و کنتراست
- نرمال‌سازی تصاویر با استفاده از میانگین و انحراف معیار دیتاست ImageNet
- تبدیل تصاویر و ماسک‌ها به Tensor

این مراحل موجب افزایش تنوع داده‌ها و بهبود تعمیم‌پذیری مدل می‌شوند.

۴. معماری مدل پیشنهادی

برای حل مسئله قطعه‌بندی پیکسلی، از معماری معروف U-Net استفاده شده است. این معماری شامل دو بخش اصلی است:

۱.۴ مسیر استخراج ویژگی (Encoder)

در این مسیر، ویژگی‌های سطح پایین تا سطح بالا استخراج می‌شوند. هر بلوک شامل:

- دو لایه Convolution
- Batch Normalization
- تابع فعال‌ساز ReLU
- لایه Max Pooling برای کاهش ابعاد

۲.۴ مسیر بازسازی (Decoder)

در مسیر بازسازی، با استفاده از لایه‌های Upsampling و اتصال‌های میان‌بر (Skip Connections)، اطلاعات مکانی از لایه‌های اولیه به لایه‌های نهایی منتقل می‌شود. این ساختار امکان بازیابی دقیق مرزها و جزئیات مکانی را فراهم می‌کند.

در انتها، یک لایه Convolution با کرنل 1×1 برای تولید نقشه قطعه‌بندی نهایی با ۳ کلاس خروجی استفاده شده است.

۵. فرایند آموزش مدل

مدل به مدت ۱۵ دوره (Epoch) آموزش داده شده است. تنظیمات آموزش به شرح زیر می‌باشد:

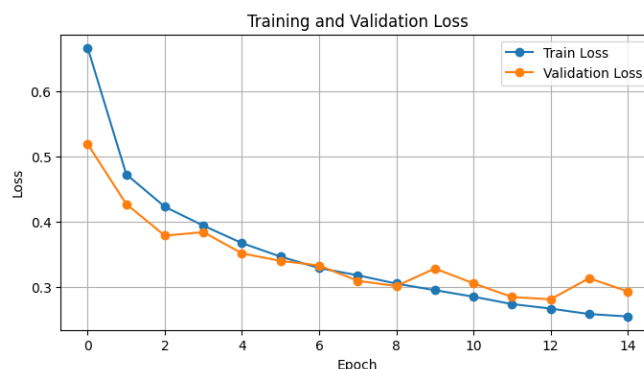
- تابع هزینه: Cross Entropy Loss
- بهینه‌ساز: Adam
- نرخ یادگیری: $1e-4$
- اندازه دسته: 4 (Batch Size)
- استفاده از GPU (CUDA) در صورت دسترسی

در طول آموزش، مقدار خطای آموزش و اعتبارسنجی و همچنین دقت پیکسلی مدل ثبت شده است.

۶. تحلیل نتایج آموزش

۱.۶ نمودار تغییرات خطا

در شکل ۲ نمودار تغییرات Train Loss و Validation Loss در طول دوره‌های آموزش نمایش داده شده است. مقدار Train Loss از ۰,۶۶۶ در ابتدای آموزش به ۰,۲۵۴ در پایان دوره پانزدهم کاهش یافته است.



شکل ۲ - نمودار تغییرات خطای آموزش و اعتبارسنجی مدل

همان‌طور که مشاهده می‌شود، خطای آموزش به‌صورت یکنواخت کاهش یافته و خطای اعتبارسنجی نیز روندی کلی کاهشی دارد. این موضوع نشان‌دهنده یادگیری مناسب مدل و عدم بیش‌برازش شدید است.

۲.۶ دقت پیکسلی (Pixel Accuracy)

دقت پیکسلی مدل در پایان آموزش به مقدار ۹۰,۲۲٪ رسیده است که نشان‌دهنده تفکیک مناسب ناحیه حیوان از پس‌زمینه می‌باشد. مقادیر مربوط به تغییرات خطا و دقت پیکسلی در برخی از دوره‌های آموزشی در جدول ۱ ارائه شده‌اند.

جدول ۱ - مقادیر Train Loss، Validation Loss و دقت پیکسلی در برخی دوره‌های آموزشی

Pixel Accuracy	Validation Loss	Train Loss	Epoch
۷۳,۶۲٪	۰,۵۱۸۴	۰,۶۶۶۱	۱
۸۵,۹۶٪	۰,۳۵۰۹	۰,۳۶۶۷	۵
۸۸,۷۸٪	۰,۳۲۷۷	۰,۲۹۴۵	۱۰
۹۰,۲۲٪	۰,۲۹۲۷	۰,۲۵۴۲	۱۵

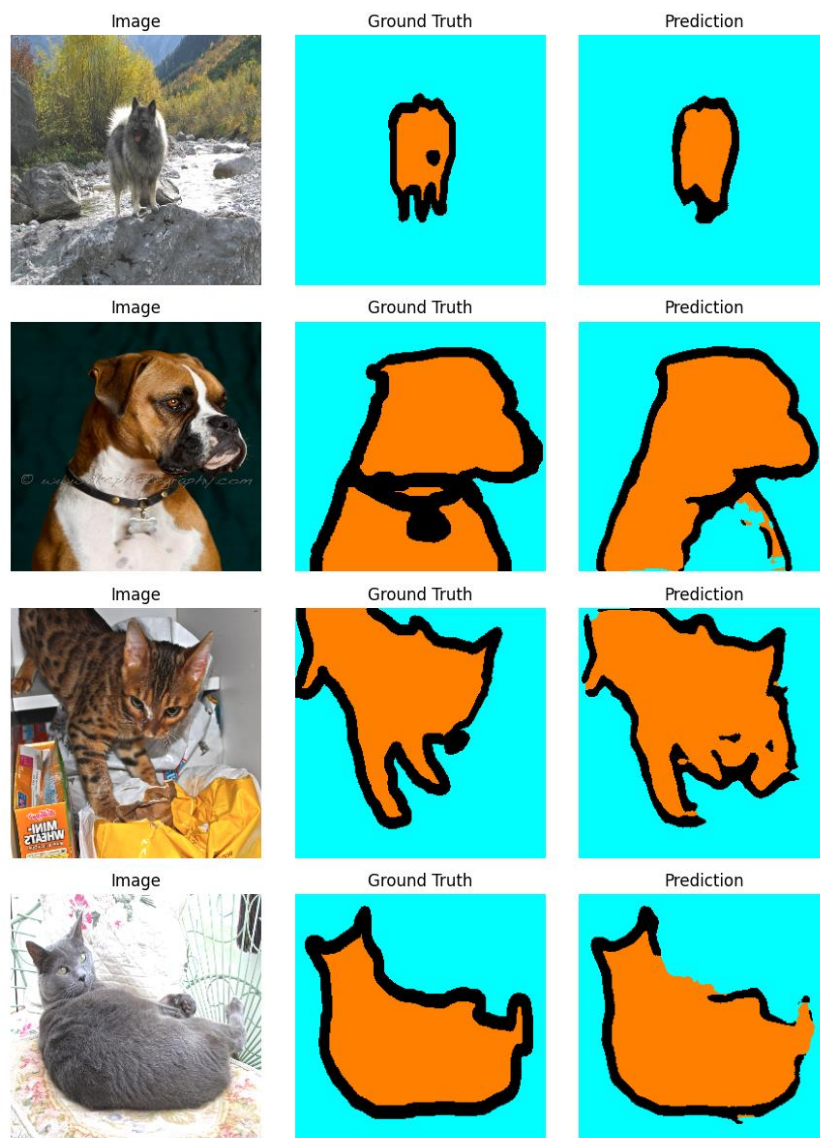
همان‌طور که در جدول ۱ مشاهده می‌شود، با افزایش تعداد دوره‌های آموزشی، مقدار خطا کاهش و دقت پیکسلی افزایش یافته است.

۷. نتایج بصری قطعه‌بندی

در شکل ۳ چند نمونه تصادفی از نتایج قطعه‌بندی مدل نمایش داده شده است. در هر ردیف:

- تصویر ورودی

- ماسک واقعی (Ground Truth)
- خروجی پیش‌بینی شده توسط مدل



شکل ۳ - نمونه‌هایی از نتایج قطعه‌بندی تصاویر با استفاده از مدل U-Net

بررسی نتایج نشان می‌دهد که مدل در اکثر موارد توانسته است ناحیه بدنه حیوان را با دقت بالا استخراج کند و مرزهای کلی شیء را به‌درستی تشخیص دهد.

۸. جمع‌بندی و نتیجه‌گیری

در این پروژه، یک مدل یادگیری عمیق مبتنی بر معماری U-Net برای قطعه‌بندی پیکسلی تصاویر حیوانات خانگی پیاده‌سازی و ارزیابی شد. نتایج به‌دست‌آمده نشان می‌دهند که:

- معماری U-Net توانایی بالایی در استخراج ساختار مکانی اشیاء دارد

- استفاده از افزایش داده و نرمال‌سازی موجب بهبود عملکرد مدل شده است
- روند کاهش خطا و افزایش دقت پیکسلی نشان‌دهنده آموزش پایدار مدل است
- نتایج بصری حاکی از تفکیک مناسب ناحیه حیوان از پس‌زمینه می‌باشند

این پروژه نمونه‌ای مناسب از کاربرد یادگیری عمیق در مسائل قطعه‌بندی تصاویر بوده و می‌تواند به‌عنوان پایه‌ای برای پروژه‌های پیشرفته‌تر در حوزه بینایی ماشین مورد استفاده قرار گیرد.