گزارش جامع پروژه مبانی بینایی کامپیوتر: اسکن و بهبود کیفیت اسناد

محمد امین حاجی علیرضایی، 40216623

۱۹ مرداد ۱۴۰۵

فهرست مطالب

بخش اول: معماری شبکه‌های عصبی (**CNN Architectures**) ۲

بخش دوم: مهندسی داده‌ها و حل چالش‌های سیستمی ۲

بخش سوم: توابع زیان و استراتژیِ تنظیم (**Regularization & Losses**) ۳

بخش چهارم: سیر تکامل مدل‌ها (**Evolution & Champions**) ۳

بخش پنجم: منطق سیستمیک، الگوریتم **Ensemble** و رابط کاربری ۵

نتیجه‌گیری ۶

این گزارش، مستندات کامل مهندسی، معماری، چالش‌ها و سیر تکامل خط‌لوله (Pipeline) طراحی‌شده برای پروژه اسکن اسناد را از فاز صفر تا استقرار در رابط کاربری نهایی (Production) شرح می‌دهد. هدف اصلی این پروژه تبدیل عکس‌های خام و تخریب‌شده‌ی گوشی‌های هوشمند به اسکن‌های تمیز، تخت و با کیفیت بالا است. این سیستم از دو وظیفه اصلی تشکیل شده است: تشخیص گوشه‌ها (Corner Detection) و بهبود کیفیت (Document Enhancement).

بخش اول: معماری شبکه‌های عصبی (CNN Architectures)

پروژه بر پایه شبکه‌های عصبی پیچشی کاملاً سفارشی (بدون استفاده از وزن‌های از پیش آموزش‌دیده) پیاده‌سازی شده است. تمامی معماری‌ها در فایل `model.py` پیاده‌سازی شده‌اند.

- مدل بهبود کیفیت (**Enhancement U-Net**): یک معماری Encoder-Decoder مبتنی بر U-Net است که شامل ارتباطات میان‌بر (Skip Connections) برای حفظ جزئیات ظریفِ متن می‌شود. این شبکه دارای حدود 31 میلیون پارامتر قابل آموزش بوده و ورودی/خروجی آن تصاویری با رزولوشن 512×512 و 3 کانال رنگی (RGB) هستند.
- تشخیص گوشه رویکرد الف (**Corner Regression**): یک شبکه استخراج ویژگی (CNN Extractor) که به یک هد تمام‌متصل (Fully Connected) ختم می‌شود. این مدل 8 عدد (مختصات x و y برای 4 گوشه) خروجی می‌دهد و حدود 14.5 میلیون پارامتر دارد. به دلیل استفاده از لایه `Flatten`، این شبکه درک فضاییِ دوبعدی را از دست داده و دقت پایینی (MLE حدود 68.2 پیکسل) ثبت کرد.
- تشخیص گوشه رویکرد ب (**Corner Heatmap**): معماری پیروز و قهرمانِ پروژه. این مدل با استفاده از بک‌بون U-Net چهار نقشه حرارتی (Heatmap) گاوسی برای هر گوشه تولید می‌کند. برای استخراج دقیق و مشتق‌پذیرِ مختصات، از لایه `SoftArgmax2D` با دمای دوگانه (Dual-Temperature) استفاده شد ($\beta = 300$ در آموزش و $\beta = 10000$ در استنتاج برای حذف اعداد اعشاری غیردقیق).

بخش دوم: مهندسی داده‌ها و حل چالش‌های سیستمی

آموزش این شبکه‌ها بر پایه "داده‌های مصنوعیِ بدون نیاز به برچسب‌گذاری دستی" انجام شد.

۱. خط‌لوله تولید داده مصنوعی (**Synthetic Pipeline**):

در کلاس `SyntheticDocumentDataset`، اسکن‌های تمیز روی پس‌زمینه‌های تصادفی قرار داده می‌شوند. 4 نقطه تصادفی روی پس‌زمینه انتخاب شده و اسکن با استفاده از ماتریس هموگرافی (Homography) و تابع `cv2.warpPerspective` روی آن‌ها نگاشت (Warp) می‌شود. این 4 نقطه به صورت خودکار همان برچسب‌های دقیقِ گوشه را تشکیل می‌دهند.

۲. تخریب‌های ساختاری و نوری:

برای شبیه‌سازی دقیق شرایط دنیای واقعی، انواع تخریب‌ها با استفاده از OpenCV در فایل `degradation.py` پیاده‌سازی شدند. این موارد شامل: تاری حرکتی و گاوسی (Motion & Gaussian Blur)، نویز سنسور، فشرده‌سازی JPEG، تغییرات روشنایی/کنتراست و دمای رنگ می‌شود.

۳. رفع چالش‌های بحرانی ؟؟؟ (**Dataset Bugs**):

- حل مشکل **Dataset Poisoning**: در ابتدا، تابع `_add_adjacent_page` به اشتباه خطوط سیاهِ ضخیمی را روی متن اصلی رسم می‌کرد. با افزودن یک عملگر ضرب داخلی (Dot Product) و بررسی بردار نرمال نسبت به مرکز سند، این مشکلِ هندسی حل شد.
- انحنای سه‌بعدی و تزریق گرافیک: برای جلوگیری از پاک شدن لوگوها و عناصر رنگی توسط مدل، تابع `_add_non_text_elements` (تزریق کلاژهای گرافیکی) و شبیه‌ساز `_apply_cylindrical_warp` (شبیه‌سازی انحنای شیرازه کتاب با تابع سینوسی) به خط‌لوله اضافه شدند.

۴. رفع قفل سیستم (**Windows IPC Deadlock**):
افزایش رزولوشن آموزش به 512×512 باعث پر شدن ظرفیت رم (OOM) و قفل شدن ارتباط بین‌پردازشی (IPC Deadlock) در سیستم‌عامل ویندوز شد. این مشکل به دلیل تداخل Thread‌های داخلی OpenCV و PyTorch DataLoader بود. راه‌حل، غیرفعال‌سازیِ کاملِ تردینگِ OpenCV با فرمان `cv2.setNumThreads(0)` و توقفِ کش کردن داده‌ها (`freeze_data=False`) در بخش آموزش بود.

بخش سوم: توابع زیان و استراتژیِ تنظیم (Regularization & Losses)

ترکیب توابع زیان بهبود کیفیت (**Enhancement Loss**):
برای جلوگیری از تولید عکس‌های تار و رنگ‌پریده، توابع زیان در `losses.py` به شدت مهندسی شدند.

- استفاده از Smooth L1 Loss با ضریب `text_weight=5.0` برای تنبیه شدیدِ مدل در صورت محو کردن جوهر تیره.
- استفاده از `Sobel Edge Loss` با وزن 2.0 برای مجبور کردن شبکه به حفظ وضوحِ لبه‌های حروف.
- استفاده از Cosine Similarity برای حفظ رنگ‌ها (`color_weight=2.0`).

کشف معمای **Dropout** (**The Dropout Fallacy**):
بررسی‌ها نشان داد اعمالِ Dropout در تمام لایه‌های شبکه‌های استخراج مختصات فضایی (مانند کرنر دتکشن) فاجعه‌بار است و منجر به "فروپاشیِ هندسی" می‌شود.

- راه‌حل (**Scheduled Bottleneck Dropout**): کلاس `DropoutScheduler` طوری اصلاح شد که تنها لایه‌های عمیق (Bottleneck) که با مقدار اولیه $p > 0.0$ تعریف شده بودند را هدف قرار دهد. با اعمال یک تاخیر 5 اپوکی (Warm-up) و استراتژیِ کسینوسی، زمان آموزش در هر اپوک متعادل شد و Dropout توانست به عنوان یک درک‌کننده معنایی (Semantic Regularizer) قدرتمند عمل کند.

بخش چهارم: سیر تکامل مدل‌ها (Evolution & Champions)

با ارزیابی مستمر، مدل‌های نهایی در سه سطح دسته‌بندی و در `models_registry` ذخیره شدند.
۱. قهرمانان تشخیص گوشه (**Corner Detection**):

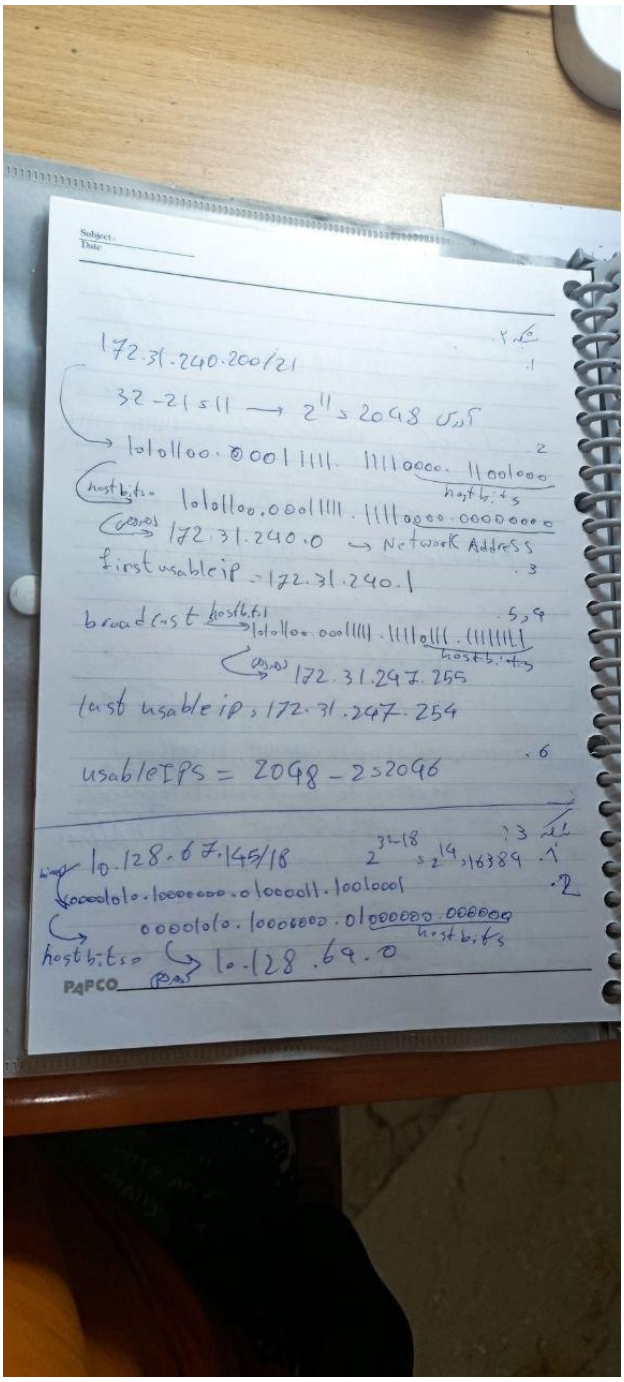
- مدال طلا (**`heatmap_v4_reg`**): آموزش دیده با Dropout زمان‌بندی‌شده در Bottleneck و داده‌های پیچیده V4. ثبت کمترین خطای ممکن روی عکس‌های واقعی (MLE برابر با 14.97 پیکسل).
- مدال نقره (**`heatmap_v3`**): بدون Dropout. یادگیری کاملاً داده‌محور (Data-Centric). خطای MLE معادل 35.04 پیکسل.
- مدال برنز (**`heatmap_v2`**): مدل پایه و ضعیف‌تر در محیط‌های شلوغ.

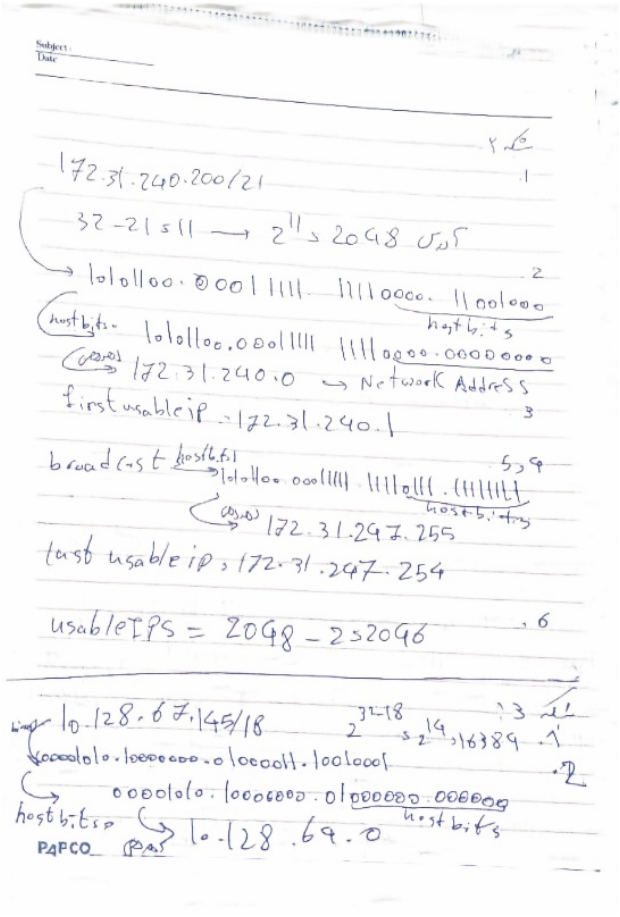
۲. قهرمانان بهبود کیفیت (**Enhancement U-Net**):
در اینجا یک تریدـآف (Trade-off) جالب بین خطای ریاضی و قابلیت خوانایی مشاهده شد:

- مدال طلا (**`enh_regularized_v2`**): بالاترین درصد موفقیت OCR (56.90%). استفاده از Bottleneck Dropout باعث شد شبکه به جای حفظ کردن پیکسل‌های نویزدار، فرم هندسی حروف را تعمیم دهد.
- مدال نقره (**`enh_clean_v2`**): بالاترین درستی پیکسلی و ریاضی با PSNR معادل 21.54 دسی‌بل و SSIM معادل 0.8544.

- مدال برنز (**enh_baseline**): مدل اولیه با باگ Dataset Poisoning (PSNR حدود 13.66 دسی‌بل).

(نکته: پارادوکس موتور *OCR* تسراکت (*Tesseract*) در عکس‌های به شدت تخریب‌شده کشف شد که به اشتباه با نمره 95% روی نویزها قفل می‌کرد. همچنین اعمال فیلتر *Gaussian Binarization* اگرچه کنتراست را مطلق می‌کرد، اما باعث افت 2.5 دسی‌بلی *PSNR* و شکستگی پیوستگی خطوط خودکار می‌شد. بنابراین *Binarization* به یک آپشن اختیاری در *UI* تبدیل شد.)





شکل ۱: نمونه اجرای عملی: تشخیص گوشه با مدل ترکیبی (Smart Ensemble) و خروجی اسکن‌شده نهایی توسط مدل قهرمان (enhancement_regularized_v2).

بخش پنجم: منطق سیستمیک، الگوریتم **Ensemble** و رابط کاربری

برای رسیدن به استانداردهای تجاری، پایپ‌لاینِ `DocumentScanningPipeline` به فیلترها و منطق‌های هوشمند مجهز شد.

۱. الگوریتم **Ensemble** هوشمند نسخه ۰.۳ (قانون تقارن پرسپکتیو):
استفاده صرف از بالاترین قطعیت (Max Confidence) منجر به "Reward Hacking" در مدل می‌شد که اشکال غیرممکنی می‌ساخت. سیستم جدید ۸۱ حالت ترکیب بین ۳ مدل طلا، نقره و برنز را با این قوانین غربال می‌کند:

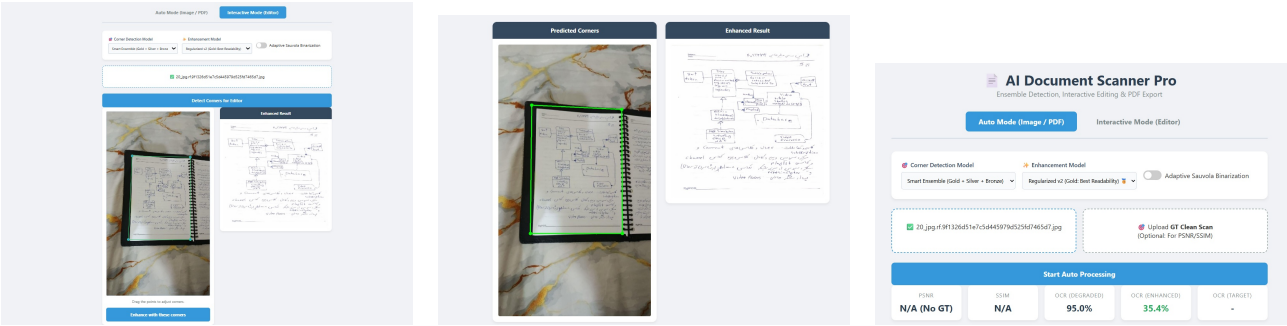
- وزن‌دهی 1.2 برابری به آرای مدل طلا.
- محدودیت زوایای داخلی بین 55 تا 125 درجه.
- قانون تقارن پرسپکتیو: اگر دو زاویه مجاور نزدیک به 90 درجه باشند، اختلاف زوایای روبه‌رو نباید بیش از 25 درجه شود.
- حداکثر نسبت اعوجاج اضلاع روبه‌رو 2.2 برابر.
- سهم 15% مساحت (Shoelace) صرفاً به عنوان گره‌گشا در صورت برابری اعتبار هندسی.

۲. دروازه‌بان‌های آماری تصاویر خارج از توزیع (**OOD Gatekeepers**):
برای جلوگیری از خراب شدن تصاویری که از قبل اسکن یا کراپ شده‌اند:

- **Gatekeeper A (Corner Bypass)**: 4 لبه تصویر (با ضخامت 3 درصد) مستقل بررسی می‌شوند. اگر حداقل 3 لبه از 4 لبه دارای واریانس پایین (زیر 800) و روشناییِ یکدست (بیش از 70% پیکسل‌ها روشن‌تر از 180) باشند، سیستم تشخیص می‌دهد عکس کراپ شده است و Corner Detection را نادیده می‌گیرد.
- **Gatekeeper C (Enhancement Bypass)**: بررسی هیستوگرامِ تصویرِ خاکستری. اگر بیش از 40 درصد از تصویر دارای روشنایی مطلقِ بالای 240 پیکسل باشد (Right-Tail Mass)، سیستم تشخیص می‌دهد این یک اسکن دیجیتال است و شبکه U-Net را برای جلوگیری از محو شدن جوهر Bypass می‌کند.

۳. رابط کاربری (**UI**) و حالت تعاملی:

- سیستم به دو حالت **Auto Mode** (اجرای خودکارِ کل خطلوله) و **Interactive Mode** (نمایش عکس در HTML5 Canvas برای ویرایش دستی گوشه‌ها با موس) تقسیم شد.
- باگِ اختلاف مقیاسِ مختصاتِ موس در Canvas از طریق ضرایب `getBoundingClientRect` رفع شد.
- امکان بارگذاری فایل‌های چندصفحه‌ای **PDF** از طریق ادغام با کتابخانه PyMuPDF افزوده شد که هر صفحه را جداگانه اسکن، کیفیت‌افزایی و سپس مجدداً در قالب یک PDF دانلودپذیر ترکیب می‌کند.



شکل ۲: تصاویری از محیط رابط کاربری (UI) سیستم در حالت‌های تشخیص خودکار و ویرایشگر تعاملی

نتیجه‌گیری

این پروژه با موفقیت از مرحله جمع‌آوری ایده به یک سیستم End-to-End قدرتمند ارتقا یافت. ترکیب داده‌های مهندسی‌شده، توابع زیان دقیق، و استراتژی‌های هندسی و آماری (Ensemble & Gatekeepers) ثابت کرد که مدل‌های CNN در کنار فیلترهای منطقی، توانایی کاملِ رقابت با اپلیکیشن‌های اسکن تجاری را دارا هستند.