hassaan research.md 2025-07-02

Landscape Study of State-of-the-Art ML Models Across Domains (2025)

1. Optical Character Recognition (OCR)

Model	Architecture	Dataset(s)	Parameters	Performance	Use Cases	Hardware
TrOCR	Transformer (Encoder- Decoder)	IAM, SROIE, STR	~330M	CER: ~2.9– 4.0%, WER: ~8%	Handwritten, printed text recognition	GPU/NPU
Tesseract	LSTM-based CRNN	Multiple public datasets	~5–100MB	CER: ~5–10%	General- purpose OCR	CPU

✓ Key Insights

• Accuracy: TrOCR > Tesseract

• Speed: Tesseract is faster on CPU, TrOCR faster on optimized GPU

• Ease of Use: Tesseract has better open-source support; TrOCR is Hugging Face-friendly

• **Use Case Match:** TrOCR excels in handwritten/scene OCR, Tesseract is better for lightweight and printed text scenarios

2. Image Classification

Model	Architecture	Dataset	Parameters	Accuracy (Top-1)	Use Cases	Hardware
EfficientNet- B7	CNN (Scaled)	ImageNet	66M	~84.3– 84.7%	General CV, mobile-friendly	GPU/CPU
ResNet-152	CNN (Residual)	ImageNet	60M	~78.3%	Legacy, wide framework support	GPU/CPU

Key Insights

• Accuracy: EfficientNet > ResNet

• **Size/Speed:** EfficientNet is more efficient with better accuracy

• Ease of Use: Both are well-documented with pretrained weights

 Use Case Match: EfficientNet is preferred for accuracy/speed tradeoff; ResNet remains a widely used baseline hassaan_research.md 2025-07-02

🕉 3. Image Segmentation

Model	Architecture	Dataset(s)	Parameters	Accuracy (mloU)	Use Cases	Hardware
DeepLabV3+	CNN (Atrous Conv)	PASCAL VOC, Cityscapes	~55–59M	~89% VOC / ~48% Cityscapes	General/medical segmentation	GPU
U-Net	CNN (Encoder- Decoder)	Custom medical/VOC	~31M	~45–85%	Medical/low- resource setups	CPU/GPU

✓ Key Insights

• **Accuracy:** DeepLabV3+ > U-Net

• Speed: U-Net is faster and lighter

• **Use Case Match:** DeepLabV3+ suits high-accuracy needs; U-Net excels in simplicity and constrained environments

***** 4. Object Detection

Model	Architecture	Dataset	Parameters	Accuracy (mAP)	Use Cases	Hardware
YOLOv8	CNN (Anchor- free)	COCO	11–68M	50–53.9%	Real-time detection	CPU/GPU
DETR	Transformer	COCO	~41M	~43–50%	End-to-end, high- quality tasks	GPU
Faster R- CNN	CNN + RPN	сосо	~41–137M	~42%	General detection	GPU

✓ Key Insights

• **Accuracy:** YOLOv8 ≥ DETR > Faster R-CNN

• **Speed:** YOLOv8 > Faster R-CNN > DETR

• Use Case Match: YOLOv8 is best for real-time; DETR suits high-fidelity pipelines

Model	Architecture	Dataset	Parameters	Accuracy (BLEU)	Use Cases	Hardware
GPT-4	Transformer	Proprietary	~1T+	~40 BLEU	Premium NLP generation	Cloud GPU

hassaan_research.md 2025-07-02

Model	Architecture	Dataset	Parameters	Accuracy (BLEU)	Use Cases	Hardware
Т5	Transformer	C4, WebText	11B	~35 BLEU	Text2Text tasks, summarization	GPU

✓ Key Insights

• Accuracy: GPT-4 > T5

• Deployment: GPT-4 requires API/cloud, T5 is open-source

• Use Case Match: T5 is flexible and local; GPT-4 provides SOTA quality at higher cost

@ 6. Multimodal Models (Vision + Language)

Model	Architecture	Dataset	Parameters	Accuracy	Use Cases	Hardware
CLIP	Vision+Text Transformer	WebImageText	~400M	76.2% Zero-shot	Retrieval, captioning	GPU
LLaVA	Multimodal Transformer	LLaVA Dataset	~13B	~58.5% VQA	Visual question answering	GPU

✓ Key Insights

• Accuracy: CLIP (zero-shot) > LLaVA (VQA tasks)

• **Speed:** CLIP is faster and lighter

• Use Case Match: CLIP for general retrieval, LLaVA for interactive QA tasks

7. Age Estimation

Model	Architecture	Dataset(s)	Parameters	Accuracy (MAE)	Use Cases	Hardware
SSR- Net	CNN (lightweight)	IMDB-WIKI, MORPH	~0.3–1M	~3.5–4.2 years	Lightweight estimation	CPU/GPU
DEX	CNN (VGG- based)	IMDB-WIKI	~138M	~3.2–5.1 years	High-accuracy setups	GPU

✓ Key Insights

Accuracy: DEX > SSR-NetSpeed: SSR-Net > DEX

• Use Case Match: SSR-Net is better for real-time or mobile; DEX suits accurate batch pipelines

hassaan_research.md 2025-07-02

Final Recommendations

Domain	Best Model	Reason
OCR	TrOCR	Highest accuracy and versatility
Image Classification	EfficientNet-B7	Best balance of size, speed, and accuracy
Image Segmentation	DeepLabV3+	SOTA semantic segmentation performance
Object Detection YOLOv8		Fast and accurate for real-time tasks
Text Generation	T5	Open-source flexibility and good performance
Multimodal CLIP		Versatile, open-source, fast
Age Estimation SSR-Net		Lightweight and efficient