


Model	Architecture	Key metric	Accuracy	Pros	Cons
FaceNet (Google)	CNN + triplet loss	Cosine similarity	99.63%	Robust model with small embeddings (128 dim)	Significant resource consumption during training
DeepFace (Facebook)	CNN (9 layers)	Euclidean + cosine	97.35%	At the time, the first model that provided human-level recognition	Hard to optimise and deploy
VGG-Face	CNN (37 layers)	Euclidean	98.95%	Deep architecture which improves accuracy	A lot of parameters, slow inference (not optimised for real-time recognition), process of recognition takes a long time
ArcFace 	CNN + Arc margin loss	Cosine similarity	99.40%	Improved spotting of differences between classes	Requires specialised and fine-tuned models
YOLO-Face	YOLO CNN	IoU (intersection over union)	99.8%	Fast all-in-one detection and recognition	Weaker performance on complex faces
Our system (CLIP + FAISS)	Clip transformer + FAISS search	L2/Cosine	~87-100% depending on size and quality of dataset	Fast, scalable, flexible, does not require training	Significant impact of embeddings' quality on performance