{% include header.html %}

{% include ai-share.html %}

文章目錄

CV

Computer Vision (電腦視覺)

AnomalyDetection

Anomaly Detection，異常檢測

- 2025-07-16：CostFilter-AD：Enhancing Anomaly Detection through Matching Cost Filtering；刷新無監督異常檢測上限！ CostFilter-AD：首個即插即用的代價濾波for異常檢測範式
- 2025-06-13：One-to-Normal：Anomaly Personalization；少樣本異常辨識新突破，擴散模型協助精準偵測
- 2025-06-06：CVPR2025, *DualAnoDiff*：Dual-Interrelated Diffusion Model for Few-Shot Anomaly Image Generation；以大模型檢測工業品異常，復旦騰訊優圖新演算法入選CVPR 2025
- 2025-05-15：**AdaptCLIP**: Adapting CLIP for Universal Visual Anomaly Detection；Github；騰訊開源 AdaptCLIP 模型刷新多領域SOTA
- 2025-05-05：Detect, Classify, Act: Categorizing Industrial Anomalies with Multi-Modal Large Language Models；DeepWiki；數據集
- 2025-04-27：**AnomalyCLIP**: Object-agnostic Prompt Learning for Zero-shot Anomaly Detection；DeepWiki
- 2025-04-26：PaDim；DeepWiki
- 2025-04-12：Anomaly-Aware CLIP, **AA-CLIP**: Enhancing Zero-shot Anomaly Detection via Anomaly-Aware CLIP；DeepWiki
- 2025-03-25：**Dinomaly**：The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection；無監督異常檢測（Unsupervised Anomaly Detection，UAD）

ObjectDetection

Object Detection (目標偵測)

- AAAI2025, Multi-clue Consistency Learning to Bridge Gaps Between General and Oriented Object in Semi-supervised Detection；Github；AAAI2025 一個遙感半監督目標偵測（半監督旋轉目標偵測）方法
- 2025-07-24：OV-DINO；開源工業開放詞彙目標偵測

- 2025-06-18：CountVid: Open-World Object Counting in Videos；牛津大學開源類別無關的影片目標計數，影片中也能「指哪數哪」
- 2025-06-15：GeoPix；像素級遙感多模態大模型
- 2025-05-23：VisionReasoner；偵測、分割、計數、問答全拿下？對標Qwen2.5-VL！ VisionReasoner用強化學習統一視覺感知與推理
- 2025-03-14：Falcon: A Remote Sensing Vision-Language Foundation Model；DeepWiki

Segmentation

Segmentation (圖像分割)

- Perceive Anything Model：Recognize, Explain, Caption, and Segment Anything in Images and Videos；對標SAM2 + LLM融合版！港中文開源感知一切模型與百萬級影像描述資料集：辨識、解釋、描述、分割一體化輸出
- RemoteSAM：Towards Segment Anything for Earth Observation
- InstructSAM：A Training-Free Framework for Instruction-Oriented Remote Sensing Object Recognition；DeepWiki
- RESAnything: Attribute Prompting for Arbitrary Referring Segmentation；Project
- CVPR 2025, Segment Any Motion in Videos, Segment Any Motion in Videos；Github
- CVPR 2025 Highlight, Exact: Exploring Space-Time Perceptive Clues for Weakly Supervised Satellite Image Time Series Semantic Segmentation；Github；Exact：基於遙感影像時間序列弱監督學習的作物提取方法
- MatAnyone：視訊摳圖MatAnyone來了，一次指定全程追踪，髮絲級還原
- Meta Segment Anything Model 2 (SAM 2)
    - 60行程式碼訓練/微調Segment Anything 2
    - CLIPSeg：Image Segmentation Using Text and Image Prompts：Huggingface Space
        - 哥廷根大學提出CLIPSeg，能同時作三個分割任務的模型
        - SAM與CLIP強強聯手，實現22000類的分割與識別
- SAMURAI
    - 無需訓練或微調即可得到穩定、準確的追蹤效果！ KF + SAM2 解決快速移動或自遮擋的物件追蹤問題
    - 經典卡爾曼濾波器改進影片版「分割一切」，網友：好優雅的方法
- Grounded SAM 2: Ground and Track Anything in Videos
    - Grounded-Segment-Anything
- SAM2Long：大幅提升SAM 2性能！港中文提出SAM2Long，複雜長視頻的分割模型
- SAM2-Adapter：SAM 2無法分割一切？ SAM2-Adapter：首次讓SAM 2在下游任務適應調校！
- SAM2Point：可提示3D 分割研究里程碑！ SAM2Point：SAM2加持可泛化任3D場景、任意提示！

- Optical Character Recognition，光學文字識別

OCR

Optical Character Recognition (光學文字識別)
針對物件或場景影像進行分析與偵測

- 2025-08-18：DianJin-OCR-R1；點金OCR-R1，模糊蓋章、跨頁表格、文字幻覺全拿下！
- 2025-07-30：dots.ocr；本地部署1.7B参数超强OCR大模型dots.ocr
- 2025-06-16：OCRFlux；DEMO；OCRFlux：一個基於LLM的複雜佈局與跨頁合併的PDF文檔解析
- 2025-06-05：MonkeyOCR；Document Parsing with a Structure-Recognition-Relation Triplet Paradigm
- 2025-05-21：PaddleOCR 3.0；OCR精準度躍升13%，支援多語種、手寫體與高精準度文件解析
- 2025-03-05：PP-DocBee：百度推出文件影像理解PP-DocBee
- 2025-03-03：olmocr：🚀本地部署最强OCR大模型olmOCR！支持结构化精准提取复杂PDF文件内容！
- 2025-02-05：MinerU：將PDF轉換為機器可讀格式的神器
- 2024-12-15：markitdown
- 2024-09-22：OCR2.0时代-GOT来啦！
- 2024-09-11：GOT-OCR-2.0模型开源
- 2024-08-20：萬物皆可AI化！剛開源就有12000人圍觀的OCR 掃描PDF 開源工具！還可轉換為MarkDown！
- advancedliteratemachinery/OCR/OmniParser
- 2024-10-29：Alibaba出品:OmniParser通用文檔複雜場景下OCR抽取
- RapidOCR
- 12個流行的開源免費OCR項目
- 用PaddleOCR的PPOCRLabel來微調醫療診斷書和收據
- TableStructureRec: 表格結構辨識推理庫來了：https://github.com/RapidAI/TableStructureRec

Diffusion model (擴散模型)

- 2025-05-28：視覺理解&生成大一統模型 Jodi；alphaXiv
- 2025-05-27：AnomalyAny；CVPR2025 | 突破資料瓶頸！ Stable Diffusion 協助視覺異常檢測，無需訓練即可產生真實多樣異常樣本
- 2025-05-23：HivisionIDPhotos，智慧證件照產生神器；AI證件照，摳圖、換背景、任意尺寸
- 2025-05-19：Index-AniSora；Aligning Anime Video Generation with Human Feedback；B站開源SOTA動畫影片生成模型Index-AniSora！
- 2025-04-26：Insert Anything；DeepWiki
- 2025-04-24：字節Phantom：1280x720影片生成革命！位元組Phantom模型實測：10G顯存效果不輸某靈付費版
- 2025-04-22：MAGI-1：Sand AI 創業團隊推出了全球首個自回歸影片生成大模型MAGI-1，該模型有哪些效能亮點？
- 2025-04-22：SkyReels V2：全球首個無限時長影片生成！新擴散模式引爆兆市場，電影級理解，全面開源
- 2025-04-14：FramePack：不是可靈用不起，而是FramePack更有性價比！開源專案：6G顯存跑13B模型，支援1分鐘影片產生
- 2025-04-14：fantasy-talking：解讀最新基於Wan2.1的音訊驅動數位人FantasyTalking
- 2025-04-05：SkyReels-A2；DeepWiki；SkyReels-A2：用AI重新定义视频创作的未来！

- 2025-03-10：HunyuanVideo-I2V：騰訊開源HunyuanVideo-I2V圖生視訊模型+LoRA訓練腳本，社群部署、推理實戰教學來吧
- 2025-02-25：Wan-Video：超越Sora！阿里萬相大模型正式開源！全模態、全尺寸大模型開源
- 2025-02-14：FlashVideo：來自位元組的視訊增強全新開源演算法，102秒產生1080P視頻
- 2025-01-28：Sana：[ICLR 2025 Oral] Efficient High-Resolution Image Synthesis with Linear Diffusion Transformer；比FLUX快100倍！英偉達聯手MIT、清華開源超快AI影像產生模型
- Flux
    - Flux.1-canny-dev：https://huggingface.co/black-forest-labs/FLUX.1-Canny-dev/
    - Flux.1-depth-dev：https://huggingface.co/black-forest-labs/FLUX.1-Depth-dev/
    - Flux.1-fill-dev：https://huggingface.co/black-forest-labs/FLUX.1-Fill-dev/
    - Flux.1-redux-dev：https://huggingface.co/black-forest-labs/FLUX.1-Redux-dev/
        - 2024-11-26：Flux官方重繪+擴圖+風格參考+ControlNet
        - 2024-11-25：最新flux_fill_inpaint模型體驗。
- 2024-12-17：Leffa：Leffa：Meta AI 開源精確控制人物外觀和姿勢的圖像生成框架，在生成穿著的同時保持人物特徵
- 2024-11-29：PuLID, Pure and Lightning ID Customization via Contrastive Alignment：https://github.com/balazik/ComfyUI-PuLID-Flux
    - 2024-11-07：搞定ComfyUI-PuLID-Flux節點只要這幾步！附一鍵壓縮包
    - 2024-10-08：一文搞懂PuLID FLUX人物換臉&風格遷移
- 2024-11-26：MagicQuill：https://huggingface.co/spaces/AI4Editing/MagicQuill
    - MagicQuill，登上Huggingface趨勢榜榜首的AI P圖神器
- 2024-11-26：OOTDiffusion：https://huggingface.co/spaces/levihsu/OOTDiffusion
    - 開源AI換裝神器OOTDiffusion
- 2024-11-24：Comfyui Impact Pack
    - Comfyui 最強臉部修復工具Impact Pack
- 2024-11-05：ComfyUI OmniGen @ 北京人工智慧研究院：https://huggingface.co/spaces/Shitao/OmniGen
    - ComfyUI 影像生成模型OmniGen，人物一致性處理的也太好了
    - 全能影像生成模型OmniGen：告別ControlNet、ipadapter等插件，僅憑提示即可控制影像生成與編輯

Digital Human (虛擬數字人)

- HeyGem：開源數位人克隆神器
- Duix：全球首個真人數位人，開源了
- Linly-Talker：an intelligent AI system that combines large language models (LLMs) with visual models to create a novel human-AI interaction method.
- EchoMimicV2：[CVPR 2025] EchoMimicV2: Towards Striking, Simplified, and Semi-Body Human Animation
- Hallo3：[CVPR 2025] Highly Dynamic and Realistic Portrait Image Animation with Diffusion Transformer Networks
- MimicTalk：[NeurIPS 2024] MimicTalk: Mimicking a personalized and expressive 3D talking face in minutes
- JoyGen：Audio-Driven 3D Depth-Aware Talking-Face Video Editing
- Latentsync
- MuseTalk