

Nguyen Ngoc Dam

Vinh Quynh, Thanh Tri, Ha Noi, Viet Nam

☎ (+84)327473388 ✉ dam.nnhust@gmail.com 🔗 <https://github.com/NNDam>

OBJECTIVE

My objective is to become an Expert AI Engineer, and through the application of my knowledge and skills, develop groundbreaking AI products capable of improving people's lives and addressing critical societal issues.

WORK EXPERIENCE

iComm Media & Tech, JSC Jun 2022 - Present

Deputy of R&D department

- Research and deploy large scale AI projects and foundation models
- Manage and support other people in team

iComm Media & Tech, JSC Aug 2018 - Jun 2022

AI Engineer

- Research and deploy AI projects

FORWARD Data Lab - UIUC (Cazoodle Inc.) Nov 2023 - Present

Research Engineer

- Develop a novel product *academic.online* (supervised by Prof. Kevin CC Chang)
- My responsibilities: Data scrapping, Data mining, NLP Engineer

SKILLS

| | |
|----------------------|--|
| Programming Language | Python, C/C++ |
| Language | Vietnamese, English |
| AI/ML/DL | <ul style="list-style-type: none">- Generative AI: LLM & Visual LLM- Computer Vision: Face eKYC (anti-spoofing, verification, identification, quality) , OCR, Intelligent Video Analytics, Image Super-Resolution / Restoration, Image Retrieval, ...- Speech Processing: Automatic Speech Recognition, Speaker Verification/Recognition, Speaker Diarization- NLP: Sentiment Analyst, Text Retrieval, NER, RAG |
| Tech Stacks | AWS, Elasticsearch, SQL, PostgreSQL, MongoDB, Milvus, FAISS, Docker, Django, FastAPI, ONNX-TensorRT, Pytorch, Kaldi, Triton Inference Server, Riva, Gstreamer, Deepstream, Kafka, RabbitMQ, CUDA, MIG, AppScript, Chrome Extension, Selenium |
| OS | Windows, Ubuntu |

HONORS & AWARDS

| | |
|---|------|
| - GOLD level in the WorldQuant Challenge | 2022 |
| - Excellent staff of the year | 2021 |
| - Excellent staff of the year | 2020 |
| - Third prize in provincial merit mathematics competition | 2015 |
| - Third prize in provincial merit Violympic competition | 2015 |

EDUCATION

PROJECTS

Multi-level Concept Generator

(06/2024 - 06/2024)

| | |
|---------------------|---|
| Customer | Academic.online |
| Description | - Generate multi-level concept keywords for concept page in product Academic.online |
| Team size | 4 |
| My position | Developer |
| My responsibilities | <ul style="list-style-type: none">- Automatic collect, scrape, parse freeform/structured HTML content from venue pages- Prompt multi LLM (Gemini, Llama 3, AWS Titan, AWS Claude 3 Haiku, ...) to extract and clean concept keywords from scrapped contexts- Post processing result to filter noisy keywords- Develop Concept generation model |
| Technologies used | - LLM, Word2Vec, Selenium |

Face eKYC & Large Scale Face Retrieval

(08/2021 - 06/2024)

| | |
|---------------------|---|
| Customer | Private (Government) |
| Description | <ul style="list-style-type: none">- Automatic/On-demand index face image from multi media sources- Provide search engine to retrieve face from large scale database- Construct relevant graph between persons |
| Team size | 5 |
| My position | Team Lead & Main Developer |
| My responsibilities | <ul style="list-style-type: none">- Research and mine face recognition data from large scale raw social, internet data- Use distributed training & reduce negative sampling to build large scale face recognition model for difference scenarios: real world surveillance system, social media, camera checkin, ... (over 6.8M IDs and 140M images)- Ranking<ul style="list-style-type: none">+ MFR-ICCV-2021 (top-5 in East Asian task, top-10 in Children task (unconstrained)): http://iccv21-mfr.com/ (NNDam)+ NIST-FRTE-11 (top 13%): https://pages.nist.gov/frvt/html/frvt11.html (datech or icomm)+ NIST-FATE-Quality (top-7): https://pages.nist.gov/frvt/html/frvt_quality.html (datech)- Develop Face Anti Spoofing module (ACER < 1%)- Optimize inference with TensorRT & Triton Inference Server- Develop indexing, quantization and distribute search algorithms for over 1B faces in 2 ways: on-disk and on-RAM (latency < 20s for top-20 on-disk and latency < 5s for on-RAM)- Develop eKYC SDK as C++ package |
| Technologies used | - Python, C++, Pytorch, Docker, Faiss, ElasticSearch, RabbitMQ, ONNX-TensorRT, Triton Inference Server, FastAPI |

IC-BCR

(05/2024 - 06/2024)

| | |
|----------|-----------------|
| Customer | Viettel Telecom |
|----------|-----------------|

| | |
|----------------------------|---|
| Description | - Automatic blur bet contents from realtime streaming |
| Team size | 5 |
| My position | Tech Lead & Developer |
| My responsibilities | <ul style="list-style-type: none"> - Develop realtime bet-contents detector using Yolo - Improve precision & recall of detector by realtime OCR - Develop pipeline to automatic blur bet content from H265-50FPS realtime streaming (video + audio) with latency ~ 1s (L4 GPU) |
| Technologies used | - Docker, Gstreamer, Deepstream, YOLO, TAO Toolkit |

Intelligent Image & Deep Video Analytics

(03/2022 - 2024)

| | |
|----------------------------|--|
| Customer | Private (Government) |
| Description | <ul style="list-style-type: none"> - Analyze image, video and stream from social media, surveillance camera and streaming for difference scenarios: retrieval, analytic, traffic optimization, management, ... - Automatic/On-demand index objects from multi media sources |
| Team size | 7 |
| My position | Researcher & Developer |
| My responsibilities | <ul style="list-style-type: none"> - Research and develop such modules using semi-supervised and domain adaption: <ul style="list-style-type: none"> + License plate detection, recognition and super-resolution + Face detection, recognition, age & gender, restoration, reconstruction, frontialization + Object segmentation and attributes recognition (person, vehicle, ...) - Research, implement and optimize video synopsis module - Optimize & develop pipeline using Deepstream and Kafka, can handle upto 60 cameras per single L4 GPU |
| Technologies used | - Python, C/C++, Elasticsearch, MongoDB, Kafka, RabbitMQ, ONNX-TensorRT, TAO, Triton Inference Server, Deepstream, Gstreamer |

IVA Assistant (in-progress)

(02/2024 - 2024)

| | |
|----------------------------|--|
| Customer | iComm Media & Tech, JSC |
| Description | <ul style="list-style-type: none"> - Assistant for Intelligent Video Analyst (IVA) system <ul style="list-style-type: none"> + Receive order/question from customer, generate plan, call defined APIs and return processed output + Generate report from given customer input (datetime, collections, ...), provide advise about security risk |
| Team size | 4 |
| My position | Developer |
| My responsibilities | <ul style="list-style-type: none"> - Develop video analyst module using Video-LLaVa & customize prompt to extract all the information about: object, context, action, reason, risk - Develop vector database and RAG module - Develop an Agent (like HuggingGPT, RestGPT) to receive question/order, plan, call system API and return output - Brainstorm future plan & features |
| Technologies used | - Python, Deepstream, Visual LLM (Video-LLaVa), RAG, Elasticsearch |

Visual Search

(03/2023 - 06/2023)

| | |
|----------------------------|--|
| Customer | Private |
| Description | - Universal objects retrieval like Google Lens, or Product Search (Amazon, Shopee, ..) but for surveillance domain |
| Team size | 5 |
| My position | Developer |
| My responsibilities | - Develop universal image embedding model using multi domain datasets & ViT-H |

| | |
|--------------------------|---|
| | pretraineds: + Finetune & evaluate different ViT-H pretrains & datasets + Ensemble 2 best models to create final embedding model + Convert & deploy model via TensorRT & Triton - Optimize & intergrade Segment-Anything model - Connect multi data sources with deep learning engine - Develop search engine with pre-filter (support on-disk & on-RAM), current database size ~ 120M |
| Technologies used | - Python, FastAPI, Faiss, ElasticSearch, RabbitMQ, ONNX-TensorRT, Triton Inference Server, OpenAI |

AI Camera

(09/2022 - 12/2022)

| | |
|----------------------------|---|
| Customer | VNVC, Tam Anh Hospital |
| Description | - Analyze stream from surveillance cameras for different tasks: + Camera checkin + Recognize loyal customers + Detect employee violations |
| Team size | 4 |
| My position | Developer |
| My responsibilities | - Research and develop AI modules using semi-supervised, domain adaption and synthetic data: garbage detection, 5S detection, employee classification, ... - Optimize & develop pipeline using Deepstream-Triton-Kafka to handle up to 20 cameras per single T4 GPU |
| Technologies used | - Python, C/C++, ElasticSearch, MongoDB, Kafka, RabbitMQ, ONNX-TensorRT, TAO, Triton Inference Server, Deepstream |

CCRating

(08/2021 - 12/2021)

| | |
|----------------------------|--|
| Customer | EVN Ha Noi |
| Description | - Automatic conversation analyst from speech audio - Automatic assessment quality of employees - Provide analysis report for hot topic weekly |
| Team size | 5 |
| My position | Researcher & Developer |
| My responsibilities | - Research and mine ASR, SID data from large scale unlabeled conversations speech data - Train & finetune ASR, SID model, diarization model for switchboard conversation domain and common domain + Transformer based (Wav2vec, Whisper) + Cascade: HMM-GMM-TDNN (kaldi), ECAPA-TDNN - Optimize and deploy models via Triton Inference Server & TensorRT - Develop automatic & on-demand workflow |
| Technologies used | - Python, Transformers, Fairseq, Kaldi, RabbitMQ, ONNX-TensorRT, Triton Inference Server, FastAPI |

ID Card Recognition

(04/2020 - 08/2020)

| | |
|----------------------------|---|
| Customer | iComm Media & Tech, JSC |
| Description | - ID Card recognition |
| Team size | 4 |
| My position | Researcher & Developer |
| My responsibilities | - Develop such modules: + Card detection + Text detection |

| | |
|--------------------------|---|
| | <ul style="list-style-type: none"> + OCR - Optimize models and develop pipeline, workflow - Develop fake-identity-detection module |
| Technologies used | - Python, Pytorch, Tensorflow, FastAPI |

Face Restoration & Recontruction & Frontialization

(06/2021 - 08/2021)

| | |
|----------------------------|--|
| Customer | iComm Media & Tech, JSC |
| Description | <ul style="list-style-type: none"> - Face restoration from low conditional environment - Reconstruct original face from some parts of face - Frontialization/normalization upto 90 degree without lossing ID information |
| Team size | 1 |
| My position | Researcher & Developer |
| My responsibilities | <ul style="list-style-type: none"> - Develop large face generation model for Asian (most Vietnam) domain - Train face restoration and reconstruction models using SOTA GAN-based method - Using face encoded model to create normalize/frontialize any unconstrained face |
| Technologies used | - Python, Pytorch, Triton Inference Server, ONNX-TensorRT, FastAPI |