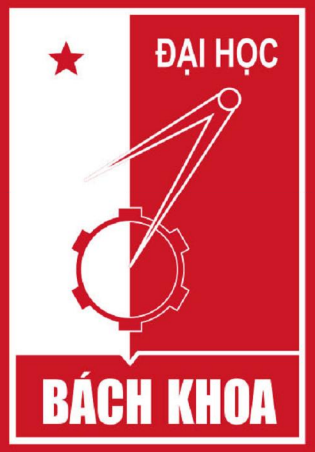


HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

HA NOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

BK.AI

–

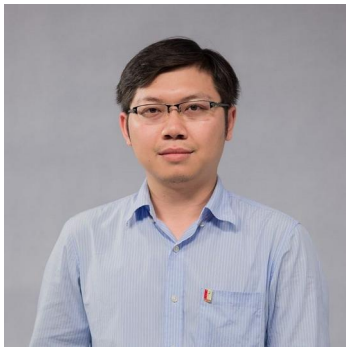
Computer Vision research group

Members

- Permanent members:
  - Dang Tuan Linh
  - Dinh Viet Sang
  - Nguyen Thi Oanh
  - Nguyen Tran Ngoc
  - Visani Muriel (leader)







- + 30 students
  - Among which 9 students regularly sit on the 10th floor

Main research partners

- **National partners (in Vietnam)**
  - USTH: ICTLab & Space department
  - MICA (HUST))
  - HUS-VNU
  - VNU-UET (FIMO)
  - VNUA (FIT)
  - HCMUS
  - Can Tho University
  - IRD: Institut de Recherche pour le Développement (Vietnam branch)

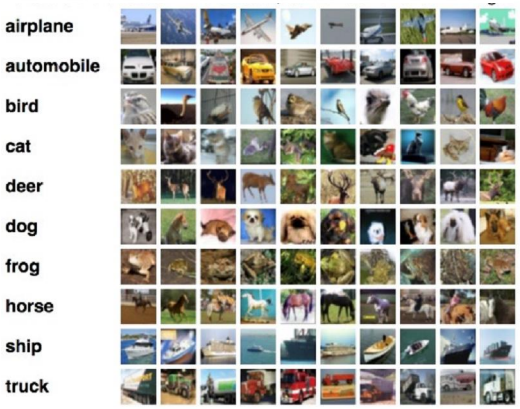
Main research partners

- **International partners**
  - **Asia-Pacific:**
    - **Australia:** University of Technology Sydney, Bureau of meteorology, CSIRO, Griffith Univ., The Univ. Of Queensland
    - **China:** Lanzhou University
    - **Japan**: Univ. of Tsukuba, Kochi Univ. of Technology
    - **Korea**: Chosun Univ.
  - **America**
    - **USA:** Univ. of Hawaii
    - **Brazil**: Univ of Sao Paulo
  - **Russia:** Tula State University
  - **Africa: Tunisia -** Sfax University (MIRACL lab)
  - **Europe**
    - **France:** La Rochelle Univ. (L3i lab), Poitiers Univ. (XLIM lab), Bordeaux Univ. (LaBRI lab), INSA Lyon (LIRIS lab), Nancy Univ. (LORIA lab)
    - **Switzerland:** Fribourg University (DIVA lab)
    - **Spain:** Barcelona University (CVC lab)

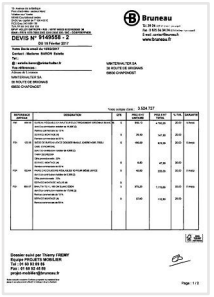
Main research Activities

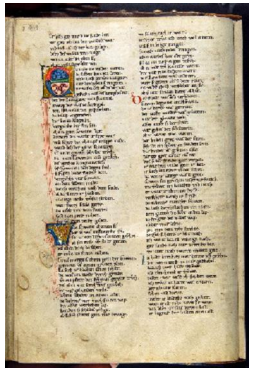
- Image/video analysis and understanding
- Different types of images / videos

– Natural scenes



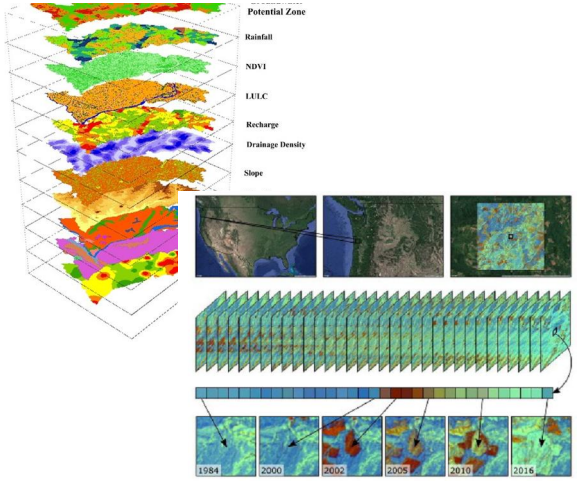
– Document images





– Medical images

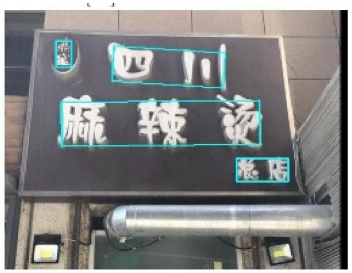
– Remote sensing

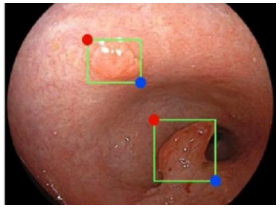


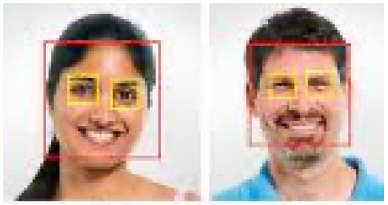
Main research Activities

- Examples of tasks
  - Without temporal information

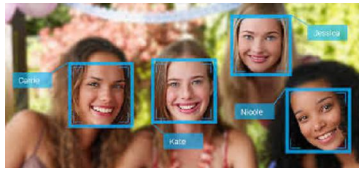
– Detection



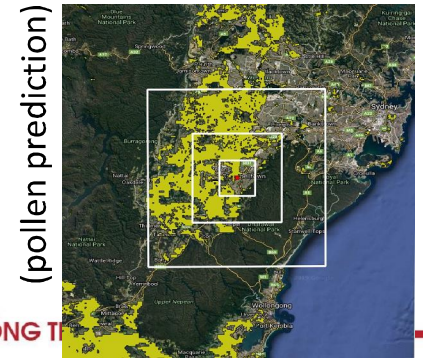




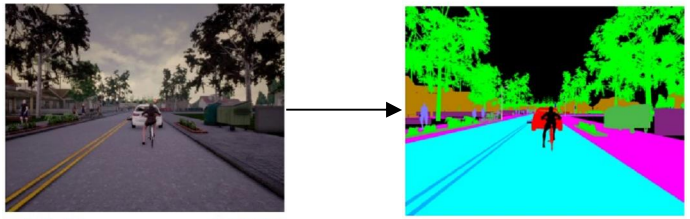
– Classification

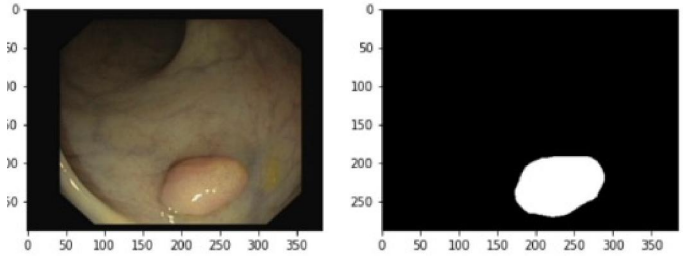


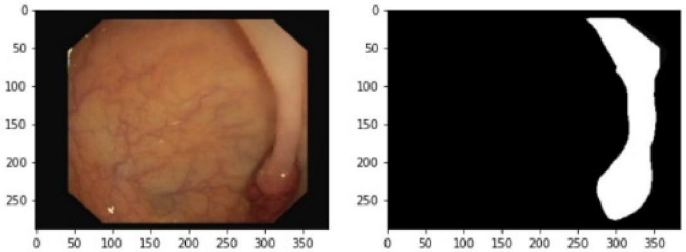
– Regression



– Semantic segmentation



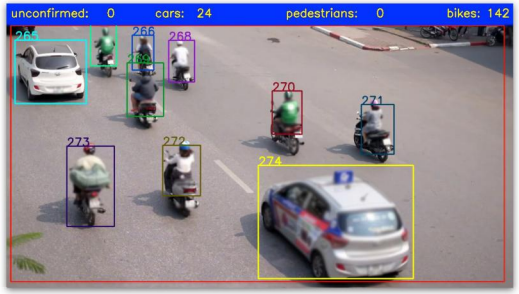




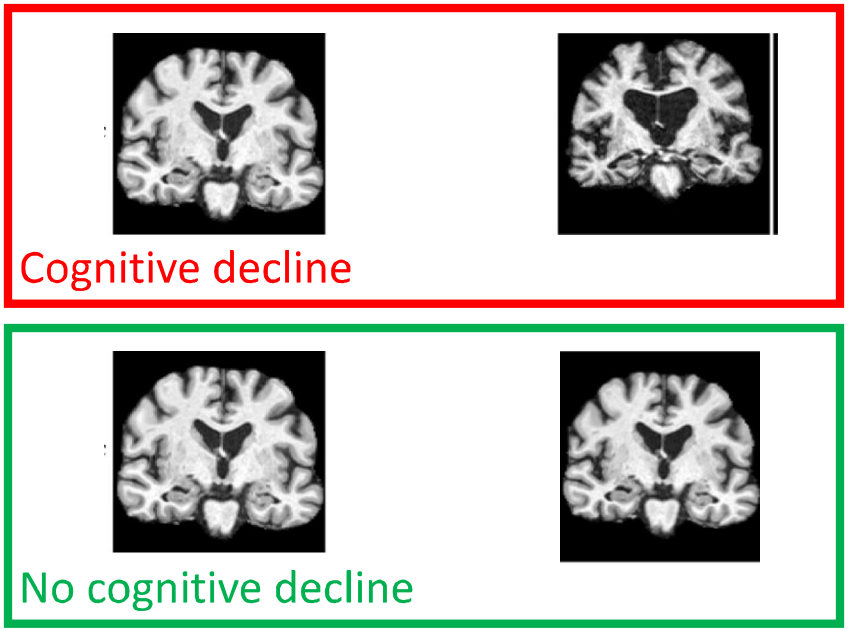
Main research Activities

- Examples of tasks
  - With temporal information

– **Tracking**
(possibly linked to embedded systems)



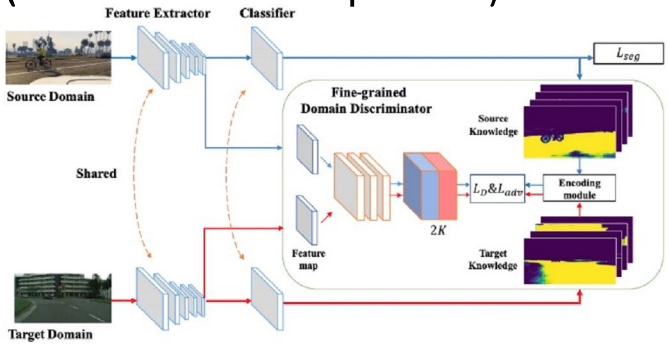
– **Detection**



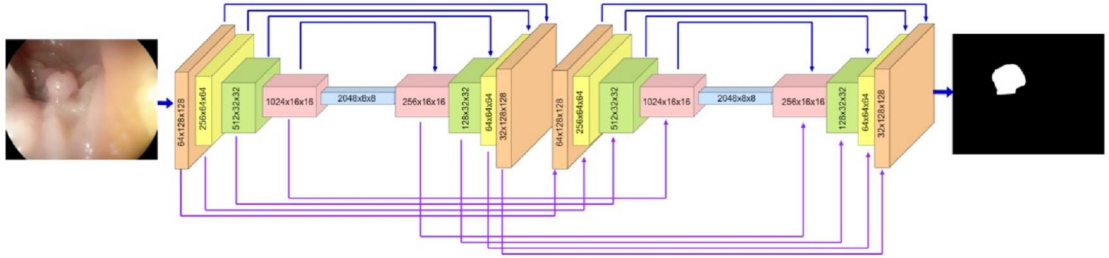
Main research Activities

- Examples of methods used
  - Image Processing methods (especially for Remote Sensing data)
  - Machine Learning (ML) methods – especially Deep Learning
    - Very often, using Convolutional Neural Networks

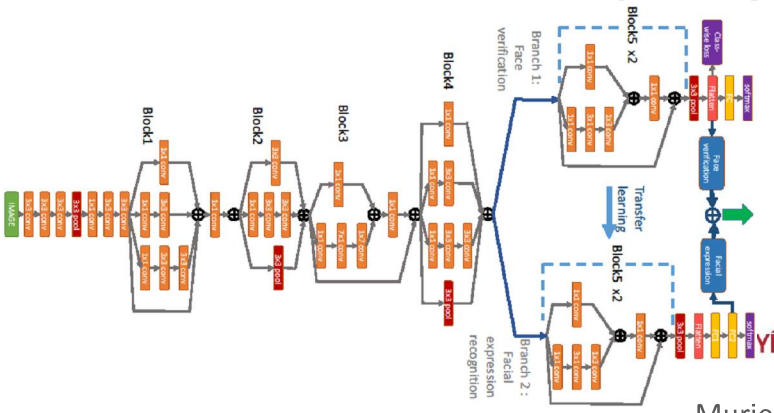
Semantic segmentation
(with domain adaptation)



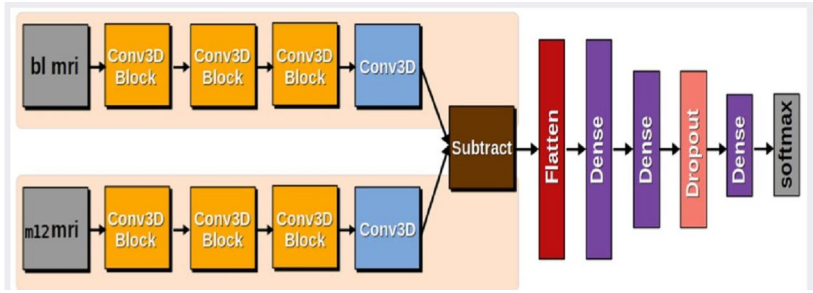
Polyp detection/segmentation from colonoscopies



Face authentication and anti-spoofing



Cognitive decline detection from 3D MRIs



Medical Imaging

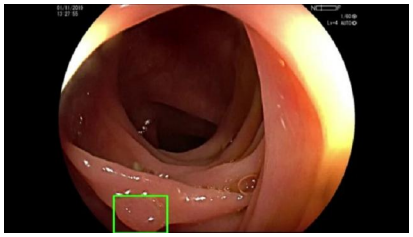
Topic 1: Segmentation of colon polyps and identifying lesions at high-risk of malignancy during endoscopy

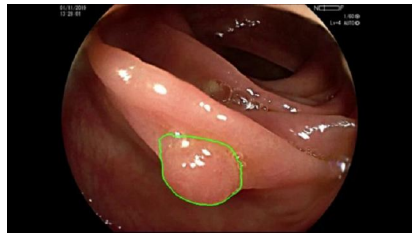
Topic 2: Detecting Brain Degeneration for Alzheimer's patients from 3D MRI images and clinical data

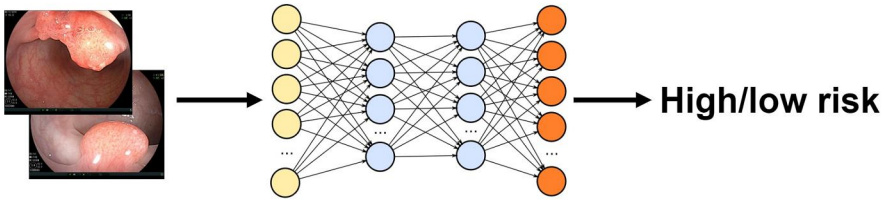
Topic 1: Detection of colon polyps and identifying lesions at high-risk of malignancy during endoscopy

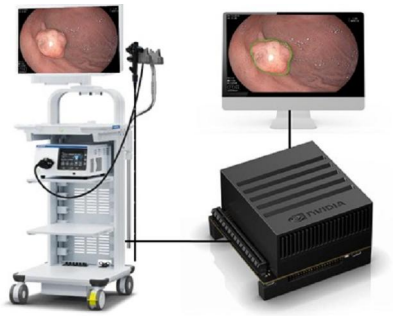
- **VinIF project (VINIF.2020.DA17)**

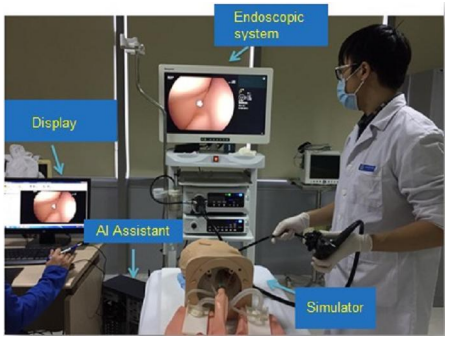
- **Goals**
  - Propose new ML models to detect/segment colon polyps and identify lesions at high-risk of malignancy during endoscopy
  - Develop a real-time system to assist endoscopists in detecting colonic polyps and identifying lesions at high-risk of malignancy





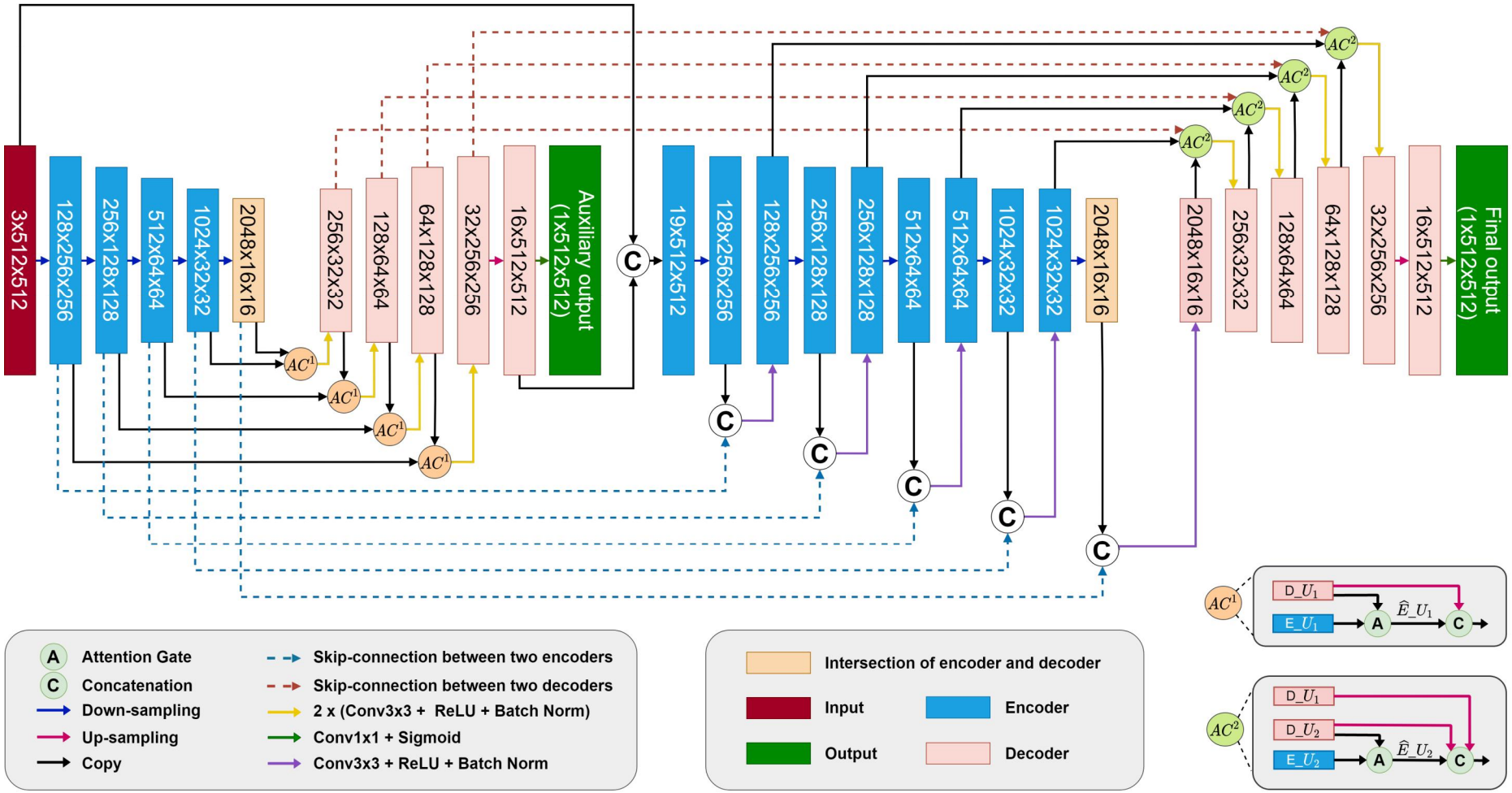






AG-CUResNeSt: Dual UNets with Attentions

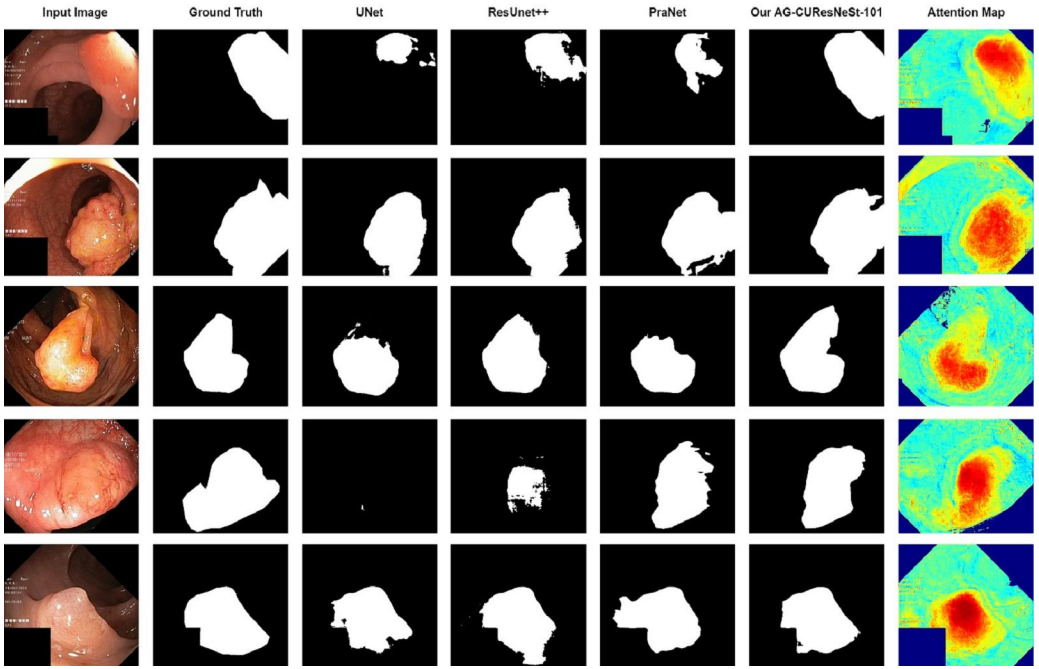
For polyp detection

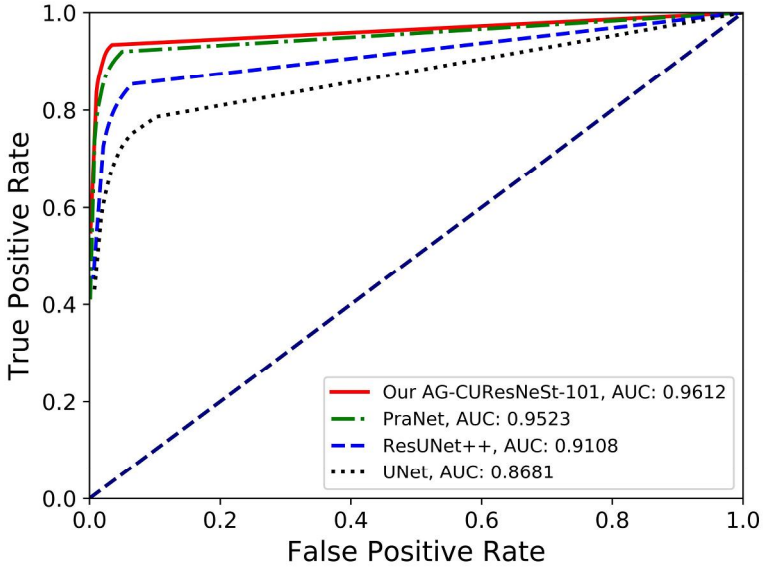


Submitted paper: https://arxiv.org/abs/2105.00402

AG-CUResNeSt: Dual UNets with Attentions

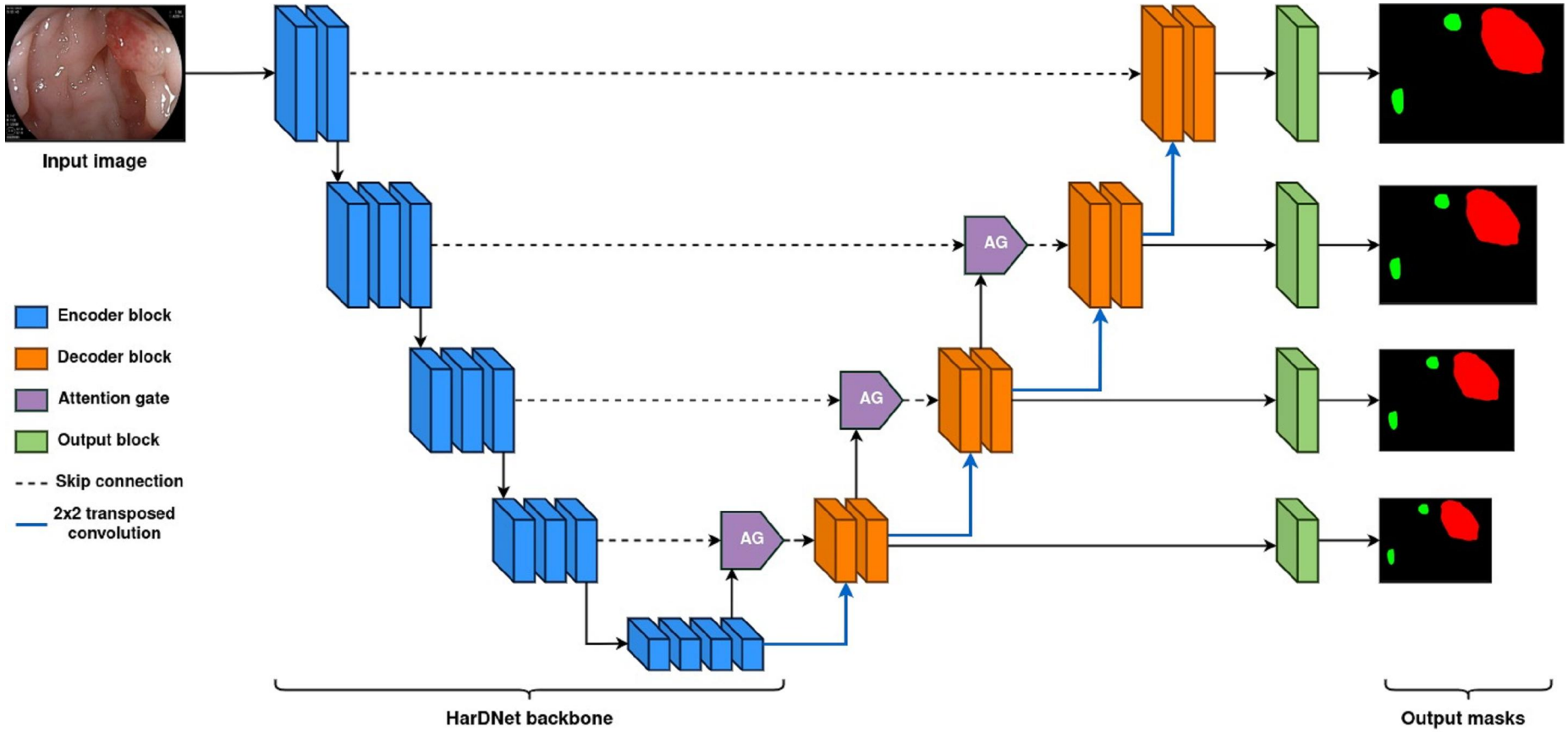
For polyp detection





NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection

*For polyp segmentation **and** detection of risk of malignancy*



[Paper submitted at PRICAI 2021]

NeoUNet: Towards accurate colon polyp segmentation and neoplasm detection

*For polyp segmentation **and** detection of risk of malignancy*

(a) Input image (b) Polyp segmentation (c) PSND

Fig. 1: Expected outputs for polyp segmentation and PSND. Black regions denote background pixels. White regions denote polyp regions. Green, red and yellow regions denote non-neoplastic, neoplastic and unknown polyp regions, respectively.

Table 1: Performance metrics on the NeoPolyp-Clean test set for U-Net, PraNet, HarDNet-MSEG, and NeoUNet

Method	$Dice_{seg}$	IoU_{seg}	$Dice_{non}$	IoU_{non}	$Dice_{neo}$	IoU_{neo}	FPS
U-Net [25]	0.785	0.646	0.525	0.356	0.773	0.631	69.6
HarDNet-MSEG [9]	0.883	0.791	0.659	0.492	0.869	0.769	**77.1**
PraNet [6]	0.895	0.811	0.705	0.544	0.873	0.775	55.6
NeoUNet	**0.911**	**0.837**	**0.720**	**0.563**	**0.889**	**0.800**	68.3

Topic 2: Predicting Brain Degeneration: context

- **Project funding:** Nouvelle-Aquitaine Region multidisciplinary research project (2017-2021) - for funding PhD students and clinical trials
  - Deep learning for 3DMRI brain image analysis
  - Partners:
    - In ICT: Poitiers University (XLIM), La Rochelle University (L3i), Bordeaux University (LaBRI)
    - Medecine doctors: from the university hospitals of Poitiers, Bordeaux and Limoges

- **Problem: detecting cognitive decline**
  - Brain tissue damage goes along with cognitive decline
  - To evaluate cognitive decline:
    - Clinical evaluations - Ex: Mini Mental State Evaluation (Folstein et. al., 1975)

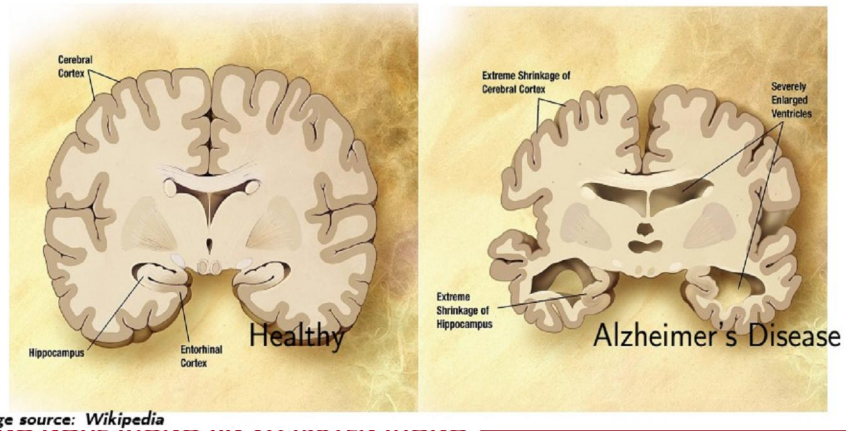


Image source: Wikipedia

MINI MENTAL STATE EXAMINATION (MMSE)

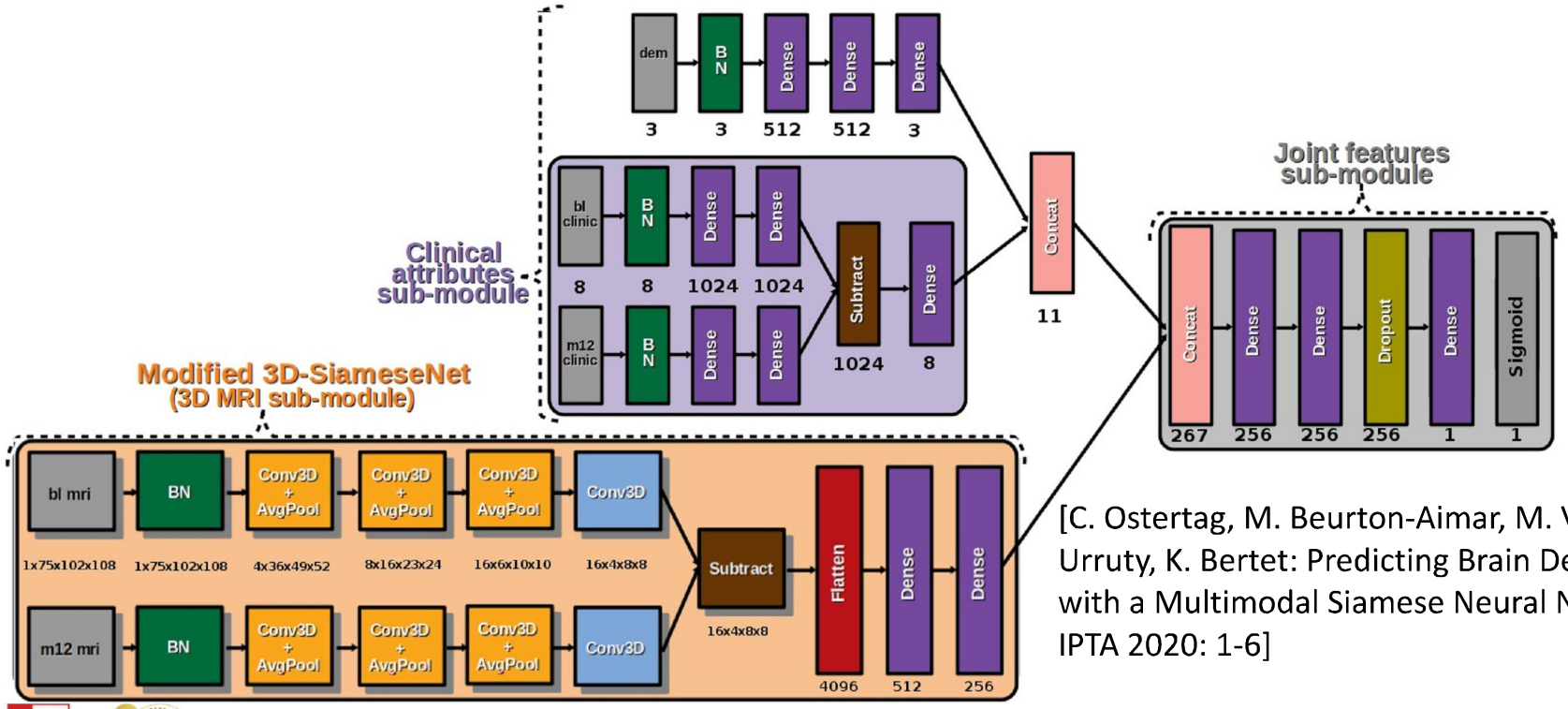
Name:
DOB:
Hospital Number:

One point for each answer	DATE:		
ORIENTATION Year Season Month Date Time	...../ 5	...../ 5	...../ 5
Country Town District Hospital Ward/Floor	...../ 5	...../ 5	...../ 5
REGISTRATION Examiner names three objects (e.g. apple, table, penny) and asks the patient to repeat (1 point for each correct. THEN the patient learns the 3 names repeating until correct).	...../ 3	...../ 3	...../ 3
ATTENTION AND CALCULATION Subtract 7 from 100, then repeat from result. Continue five times: 100, 93, 86, 79, 65. (Alternative: spell "WORLD" backwards: DLROW).	...../ 5	...../ 5	...../ 5
RECALL Ask for the names of the three objects learned earlier.	...../ 3	...../ 3	...../ 3
LANGUAGE Name two objects (e.g. pen, watch).	...../ 2	...../ 2	...../ 2
Repeat "No ifs, ands, or buts".	...../ 1	...../ 1	...../ 1
Give a three-stage command. Score 1 for each stage. (e.g. "Place index finger of right hand on your nose and then on your left ear").	...../ 3	...../ 3	...../ 3
Ask the patient to read and obey a written command on a piece of paper. The written instruction is: "Close your eyes"	...../ 1	...../ 1	...../ 1
Ask the patient to write a sentence. Score 1 if it is sensible and has a subject and a verb.	...../ 1	...../ 1	...../ 1
COPYING: Ask the patient to copy a pair of intersecting pentagons	...../ 1	...../ 1	...../ 1
TOTAL:	...../ 30	...../ 30	...../ 30

MMSE scoring
24-30: no cognitive impairment
18-23: mild cognitive impairment
0-17: severe cognitive impairment

Predicting Brain Degeneration: proposed method

- **Our work** for stable-decline classification
  - Comparing baseline *vs.* Follow-up visit
  - From 3D MRIs + clinical data (multimodal learning)
  - Using ensemble learning with **intermediate fusion** strategy



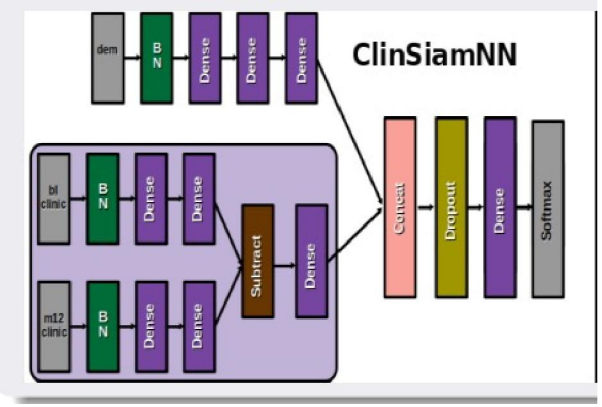
[C. Ostertag, M. Beurton-Aimar, M. Visani, T. Urruty, K. Bertet: Predicting Brain Degeneration with a Multimodal Siamese Neural Network. IPTA 2020: 1-6]

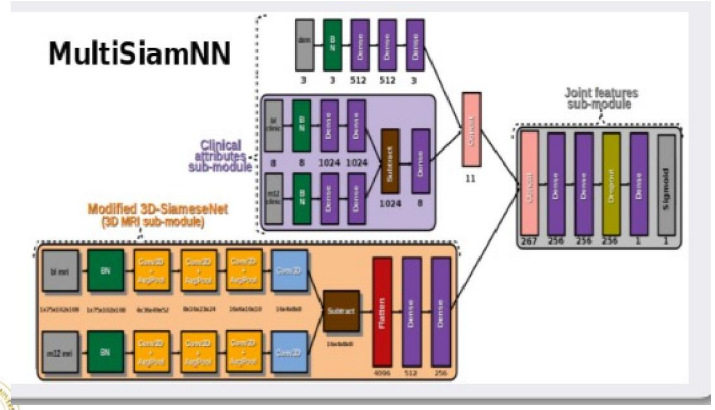
Data augmentation during training for MRIs

Predicting Brain Degeneration: results

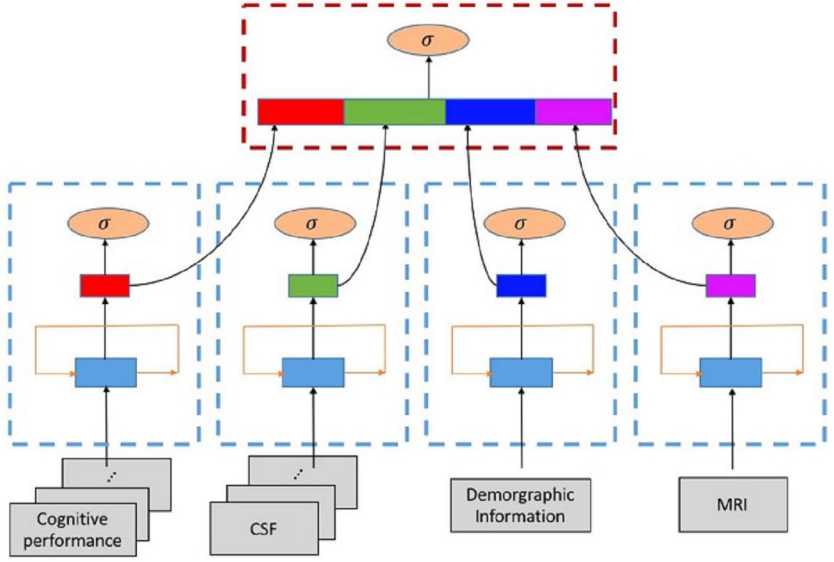
- Experimental results: models we compare

Our work





SOTA [Lee *et al.*, Scientific reports, 2019]
4 sub-Recurrent Neural Networks followed by ensemble prediction
- Sequence of medical visits for cognitive scores and CSF measures
- Baseline values for demographics and MRI measures (computed from MRIs)



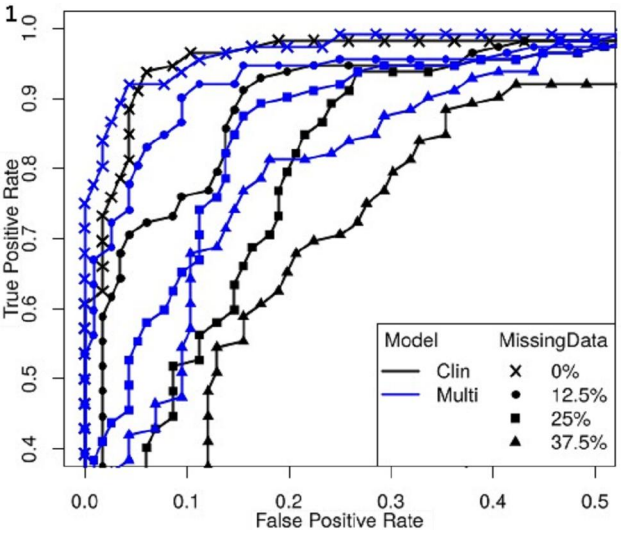
Predicting Brain Degeneration: results

- Experimental protocol
  - Dataset ADNI: 377 patients (191 Stable, 186 Decline)
  - 3 years follow-up study with regular medical visits (every 6 months).
    - Brain MRI, brain PET scans, demographic information, genetic data, biological measures, cognitive scores, . . .
  - 4-fold stratified cross-validation
  - On-the-fly data augmentation during training for MR images
- Experimental results
  - Siamese approach gives better results than RNN (SOTA), with fewer parameters
  - Our multimodal model (end-to-end, with 3D convolutions) works better than the SOTA method which uses hand-crafted features extracted from the brain image
  - Multimodal classifier is the best: shows the added value of MRI

Model	Time steps	Acc	Pre	Rec	AUC	F1
RNN *(Lee et. al.)*	bl to m12	0.846	0.807	0.840	0.918	0.822
RNN *(Lee et. al.)*	bl to m18	0.853	0.819	0.839	0.918	0.829
ClinSiamNN	bl + m06/12	0.899	0.822	0.92	0.968	0.899
MultiSiamNN	bl + m06/12	**0.925**	**0.924**	**0.929**	**0.978**	**0.925**

Predicting Brain Degeneration: results

- Dealing with missing data
  - Often a problem for medical applications
  - In this experiment, we simulate 0 to 37.5% of missing values in the Clinical Data
    - Completely at Random
  - Conclusion: MRI features (morphological) can compensate for missing clinical features (symptomatic)

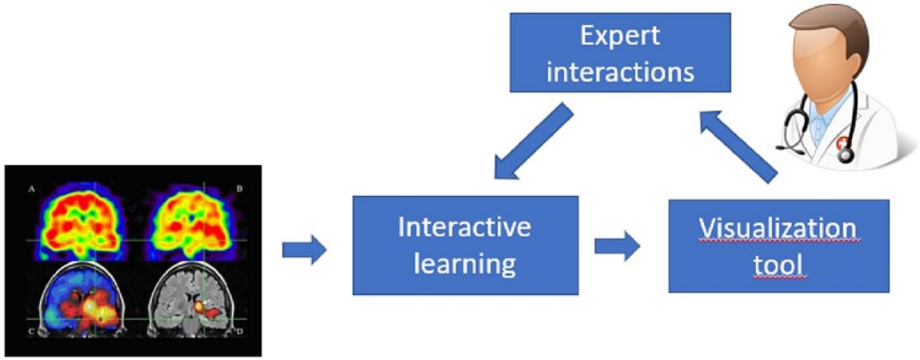


Area Under the ROC Curves comparison

MissingData	Clin	Lin.SVM	Rbf.SVM	RF	MLP	Adaboost	Multi
0%	0.968	0.860	0.856	0.867	0.863	0.865	**0.978**
12.5%	0.931	0.830	0.803	0.840	0.828	0.843	**0.951**
25%	0.876	0.770	0.742	0.783	0.769	0.786	**0.916**
37.5%	0.807	0.728	0.704	0.738	0.727	0.732	**0.873**

Predicting Brain Degeneration: conclusion

- Future work
  - Working closely with Medecine Doctors as system's Users and integrate user interaction in the Machine Learning process
    - Idea: seek for / learn from the expert user's feedback about the machine's predictions
    - Sandwich PhD (France-Brazil) starting next year
      - in the framework of another French project that I'm leading



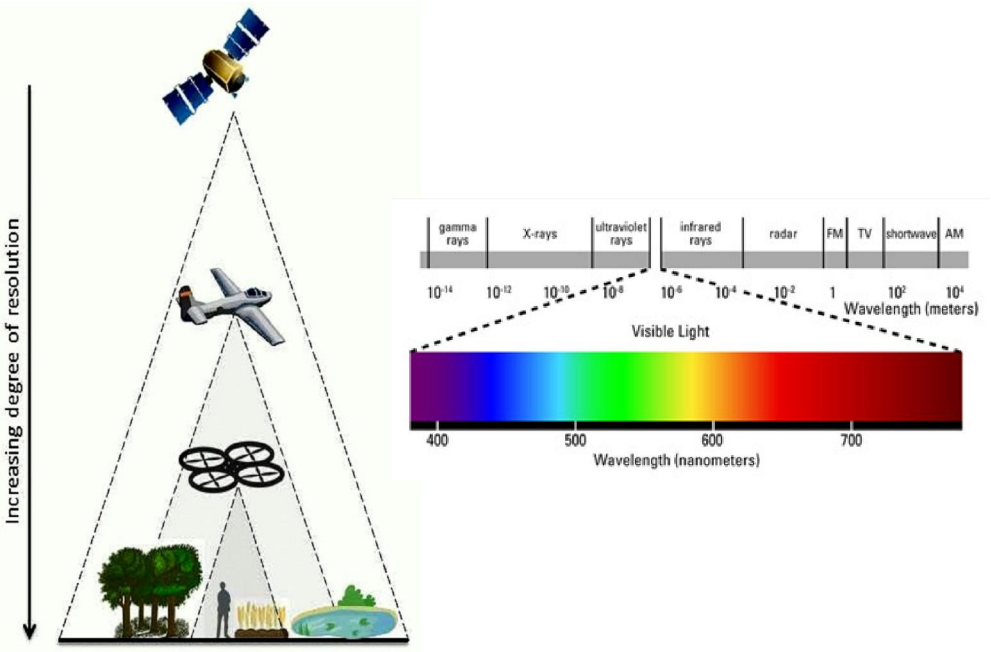
Remote sensing - satellite image processing and analysis

Topic 1: Adjusting GEO images with LEO images

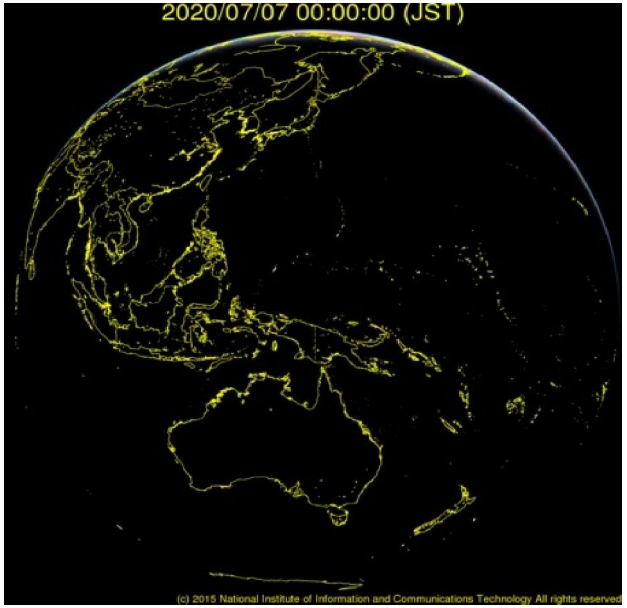
Topic 2: Urban Heat Island

Remote sensing and geostationary satellite

- Remote sensing is ***the acquisition of information*** about an object ***without making physical contact*** with the object



- Geostationary (GEO) satellite location is permanently fixed in the same position in the sky
  - Different from Low-Earth-Orbit (LEO) traditional satellites such as MODIS
  - Very useful for environmental research, especially near real-time applications

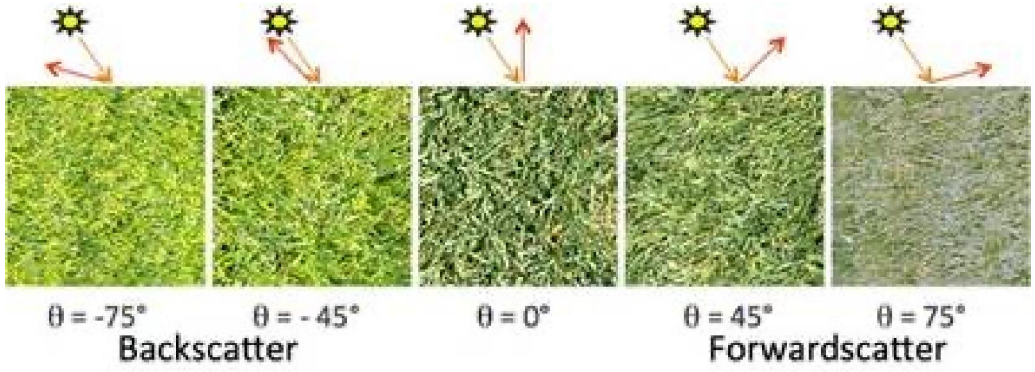


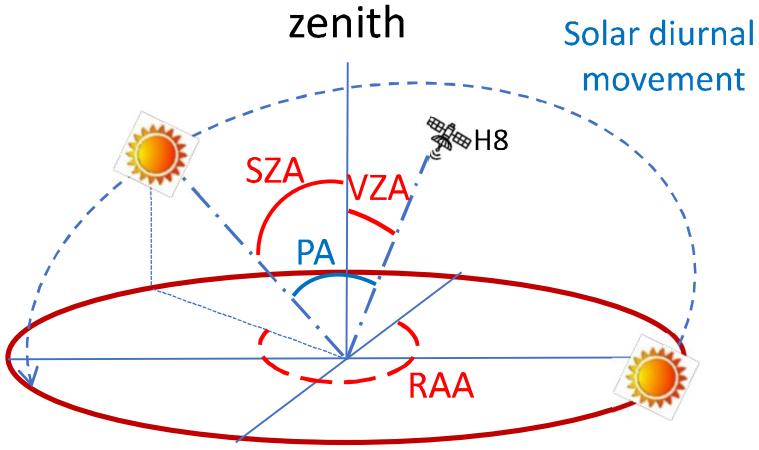
One-day Himawari-8-AHI animation (retrieved from https://himawari8.nict.go.jp)

Topic 1: Adjusting GEO images with LEO images

- Project:
  - Himawari-8 EVI development in support of TERN Phenology product and forecasting
  - Acronyms: TERN (Terrestrial Ecosystem Research Network), EVI (Enhanced Vegetation Index)
  - Period: 3 years from 2017 to 2020

Bidirectional Reflectance Distribution Function (BDRF) effect

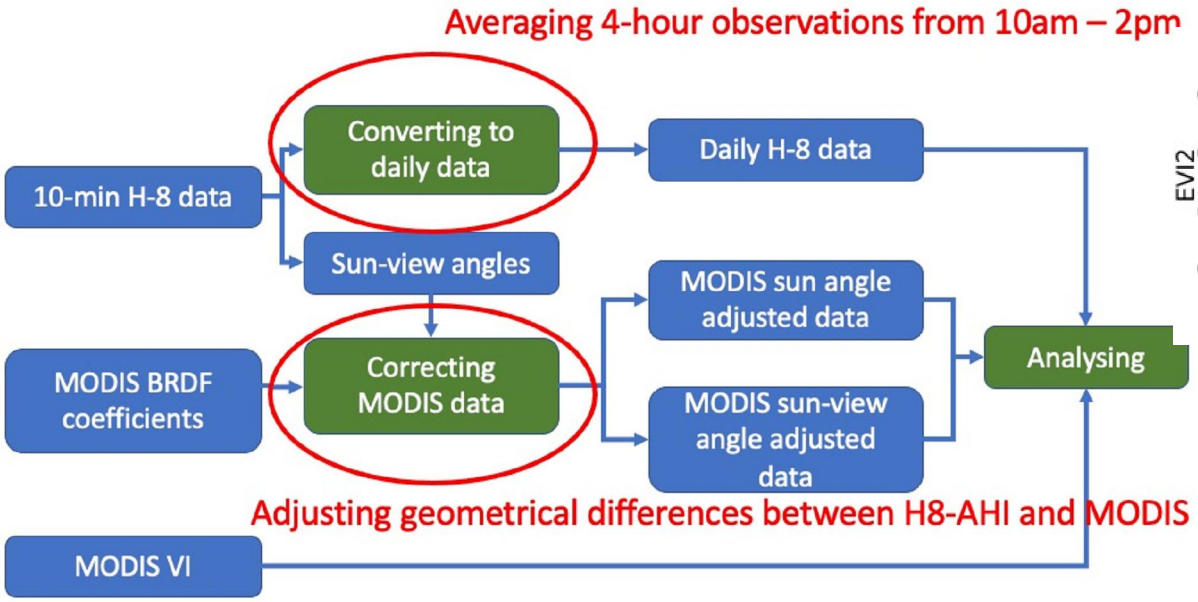


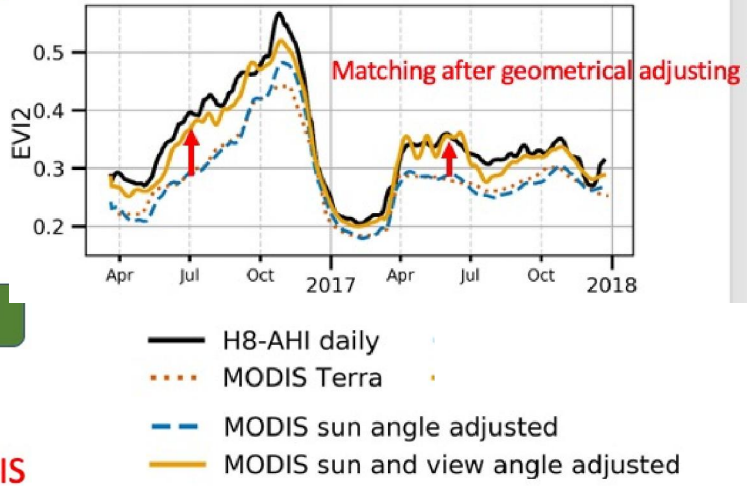


Sun zenith angle (SZA) diurnal variations

Adjusting GEO images with LEO images

- Objectives of the project
  - Generating daily H-8 data (*vegetation index*) from multiple sites in Australia
  - Matching H-8 with MODIS seasonal vegetation indices (VIs)
    - By correcting differences of sun-view geometries



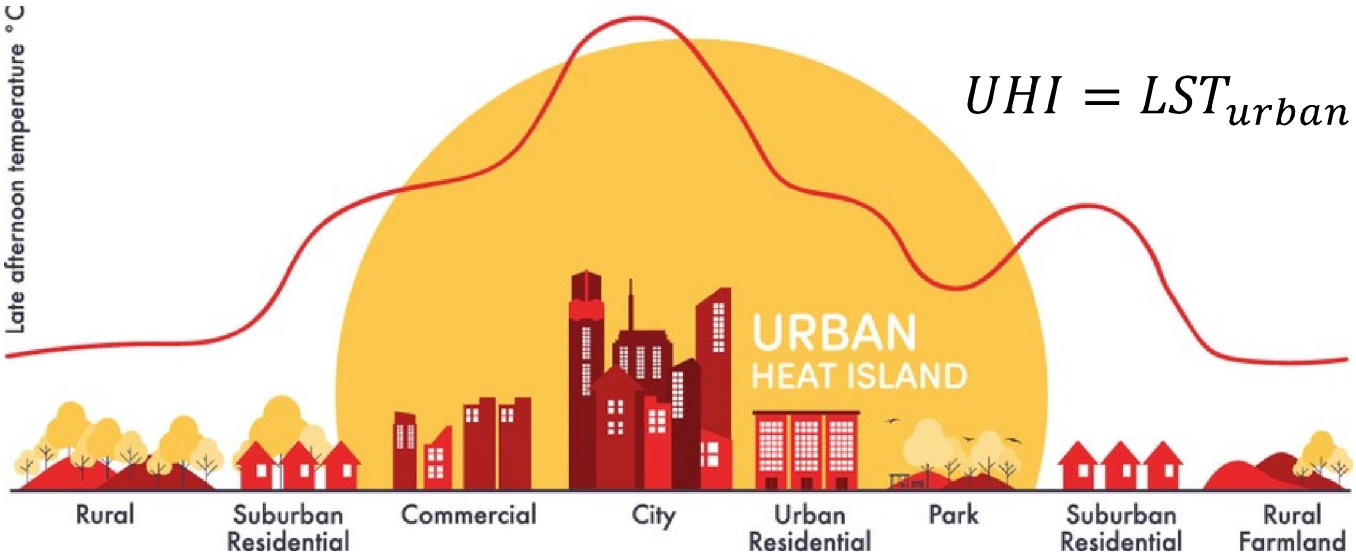


Publications related to this work and other work on H-8:

- Tran, N.N. *et al.* Seasonal Comparisons of Himawari-8 AHI and MODIS Vegetation Indices over Latitudinal Australian Grassland Sites. *Remote Sensing* **2020**, *12*, 2494.
- Ma, X.; Huete, A.; Tran, N.N.; Bi, J.; Gao, S.; Zeng, Y. Sun-Angle Effects on Remote-Sensing Phenology Observed and Modelled Using Himawari-8. *Remote Sensing.* **2020**, *12*, 1339.

Topic 2: Urban Heat Island

Project: Fairwater Living Laboratory Research Project



$$UHI = LST_{urban} - LST_{rural}$$

With LEO satellites, we can only estimate Land Surface Temperature at limited times each day

Ex: MODIS can measure LST at 10.30 am/pm and 1:30 am/pm each day

Using GEO satellites, we can detect the time of maximum LST each day

Linking LST with other factors such as hospital medical reports can be useful to understand the impact of Urban Heat Islands on human life

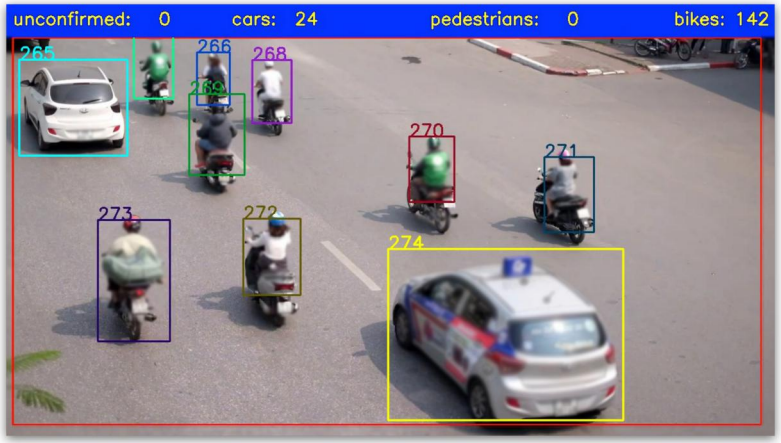
Traffic monitoring and intelligent vehicles

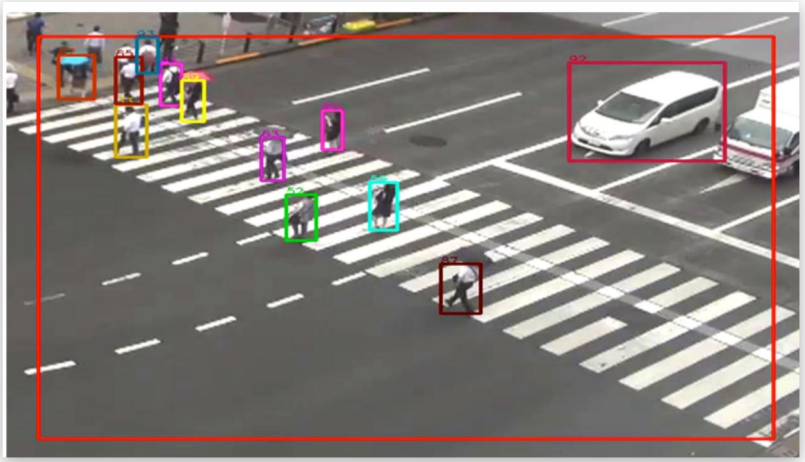
Topic 1: vehicles and pedestrian tracking in videos

Topic 2: semantic segmentation for intelligent vehicles

Topic 1: Vehicles and pedestrians tracking in video

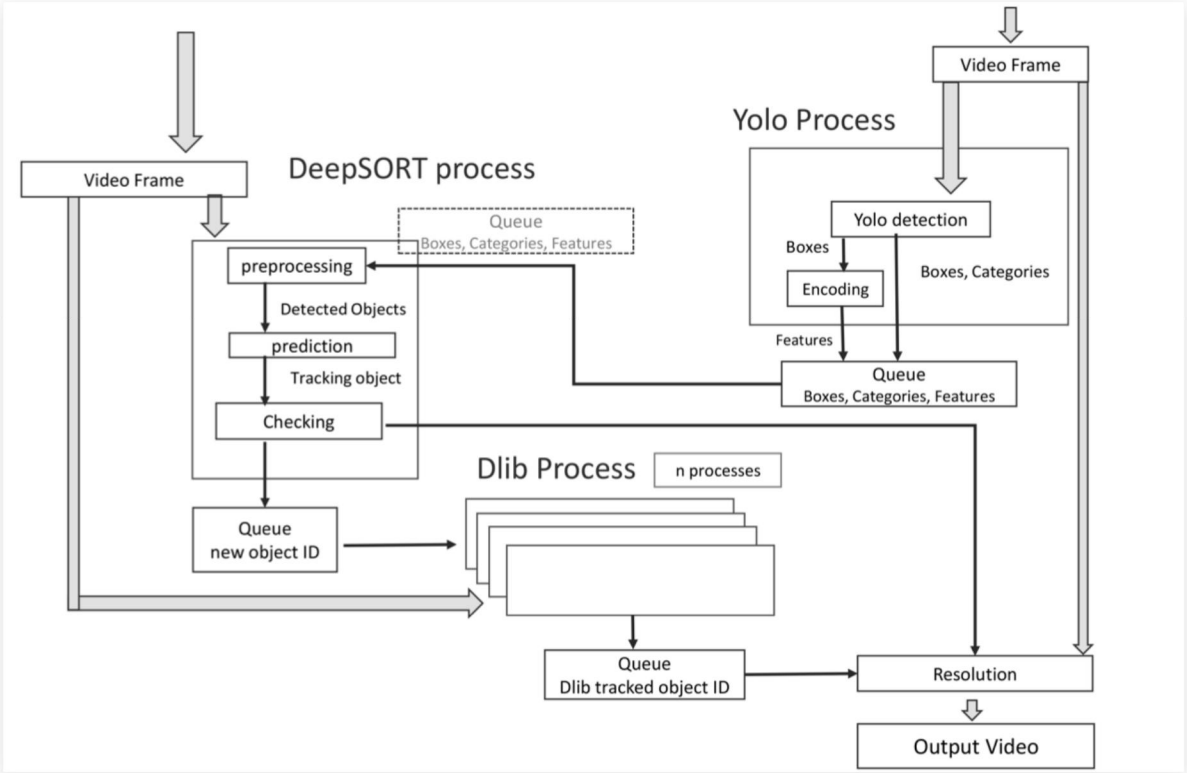
- Task: tracking vehicles and pedestrians
  - Reseach results used by a Japanese company





Vehicles and pedestrians tracking in video

- Proposed architecture
  - Reseach results used by a Japanese company



Dang, Tuan Linh, Gia Tuyen Nguyen, and Thang Cao. ”Object Tracking Using Improved Deep_Sort_YOLOv3 Architecture.” *ICIC Express Letters* 14.10 (2020): 961-969

Vehicles and pedestrians tracking in video

Number of identity switches

	Uncrowded video	Crowded video
DeepSORT-only	271/6804 (3.96%)	519/10522 (4.93%)
DeepSORT-Dlib	81/6709 (1.21%)	324/10298 (3.15%)

Operating speed

	Uncrowded video	Crowded video
Serial approach	1500 frames/195.84 seconds (7.66 FPS)	1730 frames/306.54 seconds (5.64 FPS)
Proposed parallel approach	1500 frames/121.76 seconds (12.32 FPS)	1730 frames/204.42 seconds (8.46 FPS)

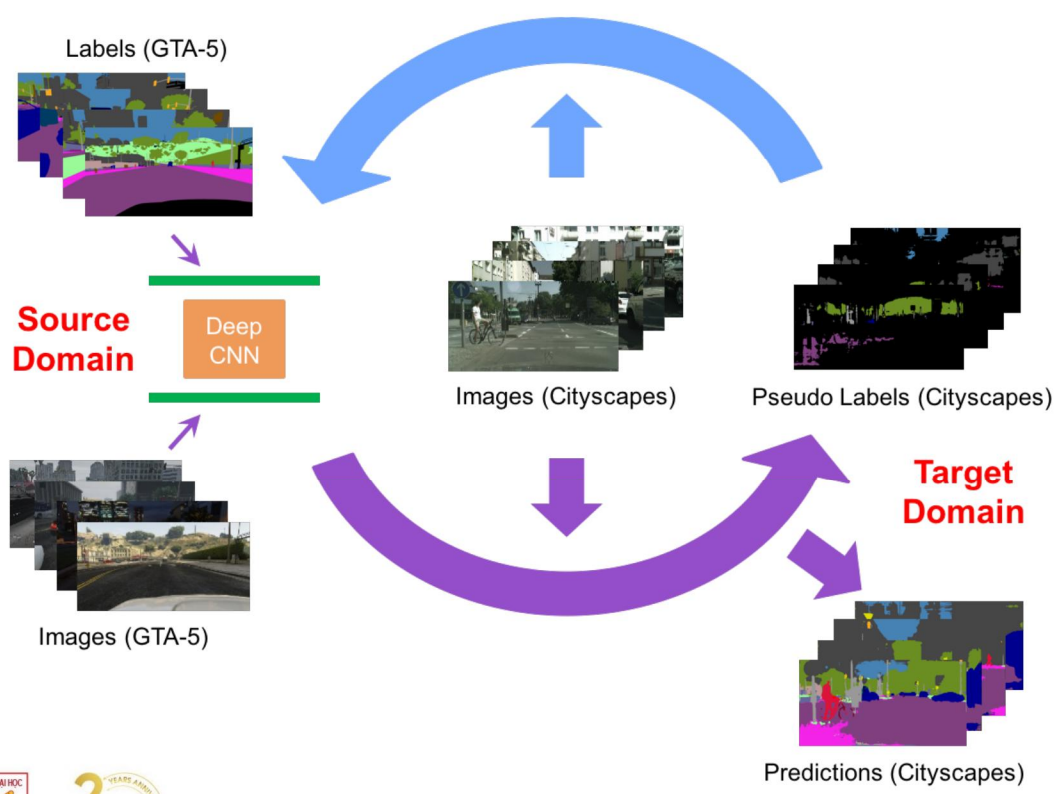
Work in progress:

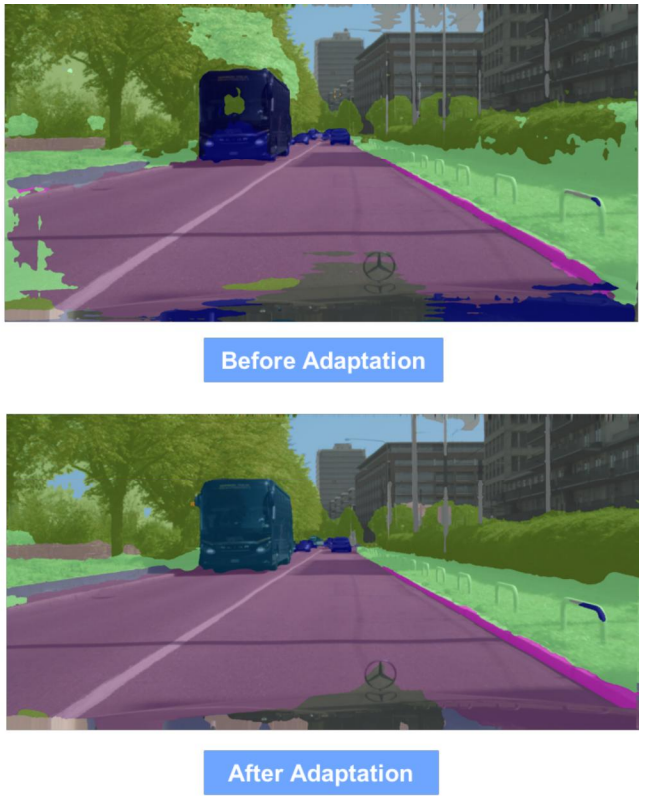
Implement and test the proposed algorithms on edge devices

- Nvidia Jetson Xavier
- Google Coral + Pi 4

Topic 2: Semantic segmentation for intelligent vehicles

- Naver project and HUST project: in progress
- Domain Adaptation and Knowledge Distillation: from simulators to real world





Gesture recognition from videos

Topic 1: Human Action Recognition

Topic 2: Hand Gesture Recognition

Topic 1: Human action recognition

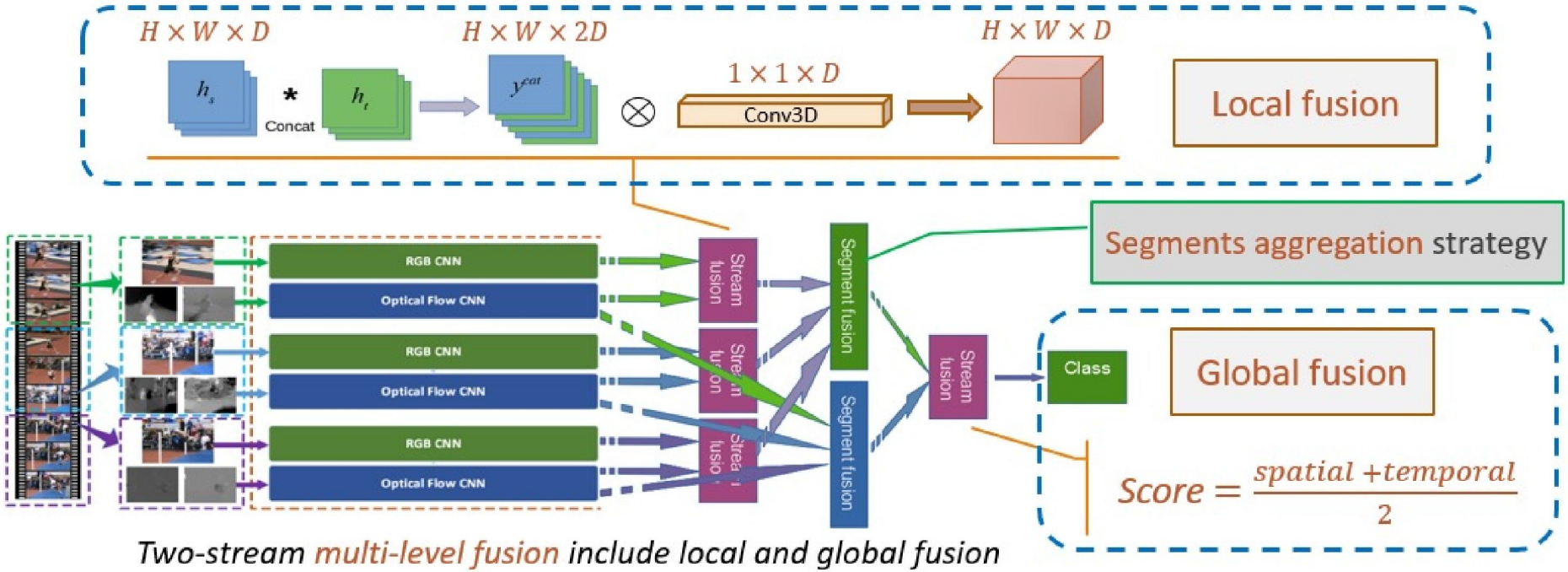
- Context:
  - Project:
    - A part of Nafosted project (2017-2020), Thai Nguyên University – Mica,HUST.
    - HUST project
- Challenges:
  - Change in appearance
    - subject, background, illumination variation, …
    - but also viewpoints
  - Occlusion

How to effectively aggregate features from different modalities or different viewpoints ??

- ***Multi-modality (spatio-temporal)***
- ***Multi-view***

Human action recognition

- Multi-modality
  - Approach: Deep learning based
    - Segments aggregation strategy to enhance temporal information : LSTM
    - Multi-level aggregation features from 2 modalities (Optical Flow+RGB)



Two-stream multi-level fusion include local and global fusion

Lu, M.H., Nguyen, T.O., 2019. Spatio-temporal Multi-level Fusion for Human Action Recognition, SoICT 2019

Human action recognition

- Multi-modality: Experimental results

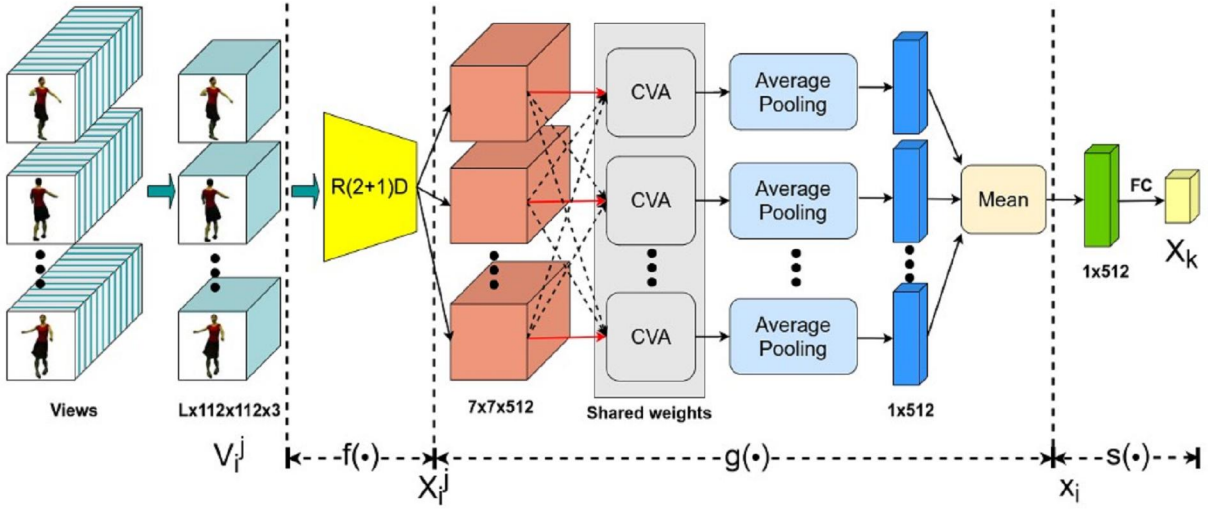
UCF-101 dataset (split 1):
- 101 classes
- Training : test 9537 : 3783

Model name	Segments aggregation	Precision	Accuracy	F1-score
Two-stream fusion at multi-level (ours)	LSTM sequence to sequence	**92.92%**	**92.57%**	**92.38%**
Global-level with early fusion		92.30%	91.17%	91.12%
Global-level with late fusion		91.95%	91.44%	91.17%

Lu, M.H., Nguyen, T.O., 2019. Spatio-temporal Multi-level Fusion for Human Action Recognition, SoICT 2019

Human action recognition

- Multi-view
  - Approaches:
    - Method 1: Enhancing Handcrafted features + SVM
    - Method 2: DL-based: multi-branches + Cross-View Attention (CVA) modules



H.-T. Nguyen, T.-O. Nguyen. Attention-based network for effective action recognition from multi-view video, KES 2021 – accepted

A.-V. Bui, T.-O. Nguyen. Multi-view Human Action Recognition Based on TSN Architecture Integrated with GRU, KES 2020 - Procedia Computer Science Volume 176, 2020, Pages 948-955.

Le, T.D., Nguyen, T.O., Tran, T.H. Improving multi-view human action recognition with spatial-temporal pooling and view shifting Techniques, SoICT'2017

Human action recognition

- Multi-view: Experimental results

Table 6: Comparison (accuracy in %) on the NUMA and MicaHandGesture datasets. V^i_{jk}: denote the videos from view j and view k are used for training and the videos from view i are used for testing. - denotes no available data.

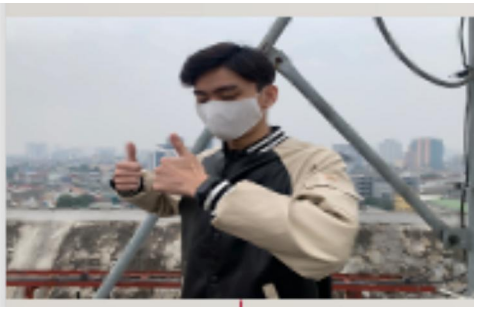
Methods	NUMA					
	Cross-subject		Cross-view			
	Video-level	Action-level	$V^1_{2,3}$	$V^2_{1,3}$	$V^3_{1,2}$	Average
MST-AOG[32]	81.6	-	-	-	-	73.3
Kong et al[14]	81.1	-	-	-	-	77.2
TSN[34]	90.3	-	76.8	80.6	84.5	80.6
DA-Net[31]	92.1	-	83.1	82.7	86.5	84.2
Glimpse Clouds [5]	-	-	83.4	89.5	90.1	87.6
Multi-Br TSN - GRU[2]	-	93.81	82.58	81.00	89.76	84.4
P-I3D[8]	-	-	-	-	90.1	-
VPN[9]	-	-	-	-	93.5	-
R34 (2+1)D With CVA	**92.78**	**99.56**	**90.47**	**91.22**	**96.52**	**92.74**

H.-T. Nguyen, T.-O. Nguyen. Attention-based network for effective action recognition from multi-view video, KES 2021 – accepted

Topic 2: Hand gestures recognition

- Naver project: in progress
- Create database ~100GB
  - 33 Static gestures
  - 27 Dynamic gestures
  - 4 cameras: top/front/left/right
  - 20 Subjects
    - 10 males
    - 10 females
    - Different backgrounds
    - Different clothes
- Study/propose models
  - Complex models
  - Lightweight models

Static hand gestures (example: “thumbs up!”)



Complex models can reach up to 96% accuracy
Lighter models can reach up to 94% accuracy

Dynamic hand gestures

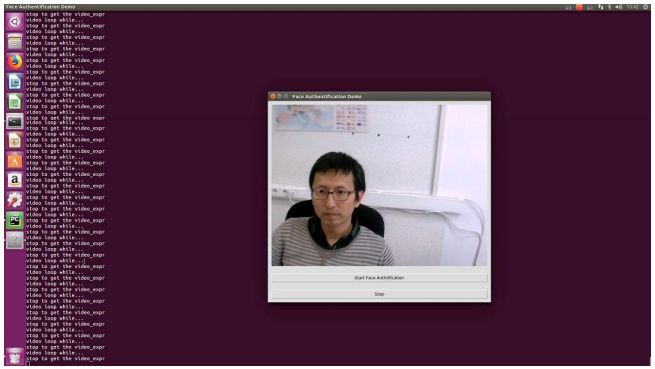
Research in progress: accuracy now ~85%

Other activities

Many other activities…

- Incremental multimodal classif. from streams of documents

- Biometry access control: face verification and anti-spoofing



- Z. Ming, M. Visani, M.M. Luqman, J.C. Burie. Journal of Imaging. 6(12):139, 56 pages, 2020
- Z. Ming, J. Chazalon, M.M. Luqman, M. Visani, J.C. Burie. Internl Conf. on Pattern Recognition (ICPR), IEEE, 2018
- Z. Ming, J. Chazalon, M.M. Luqman, M. Visani, J.C. Burie. Intl Conf. on Computer Vision Workshops, IEEE, 2017

- Understanding ancient text (Han-Nom characters) from historical monuments

- Recognize sequence of digits engraved on metal plates

- Dang, Tuan Linh, Thang Cao, and Yukinobu Hoshino. "Engraved digit detection using HOG-real AdaBoost and deep neural network." *Turkish Journal of Electrical Engineering & Computer.*

Summary/discussion

Summary: main scientific problems

- Some keywords about our scientific problems (non-exhaustive list)
  - Multimodality
    - Spatio-temporal information
    - Raw data (or text) associated to the images
  - Domain adaptation
    - Transferring the model learnt from one set of images to a different set of images
  - Limited resources constraints (linked to embedded systems)
    - Definition of light weight models
  - User interaction
  - …

ĐẠI HỌC BÁCH KHOA

25 YEARS ANNIVERSARY SOICT

VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY

Thank you for your attention!

soict.hust.edu.vn/ fb.com/groups/soict

Questions