Stony Brook University

Computer Vision and Applications in the Deep Learning Era

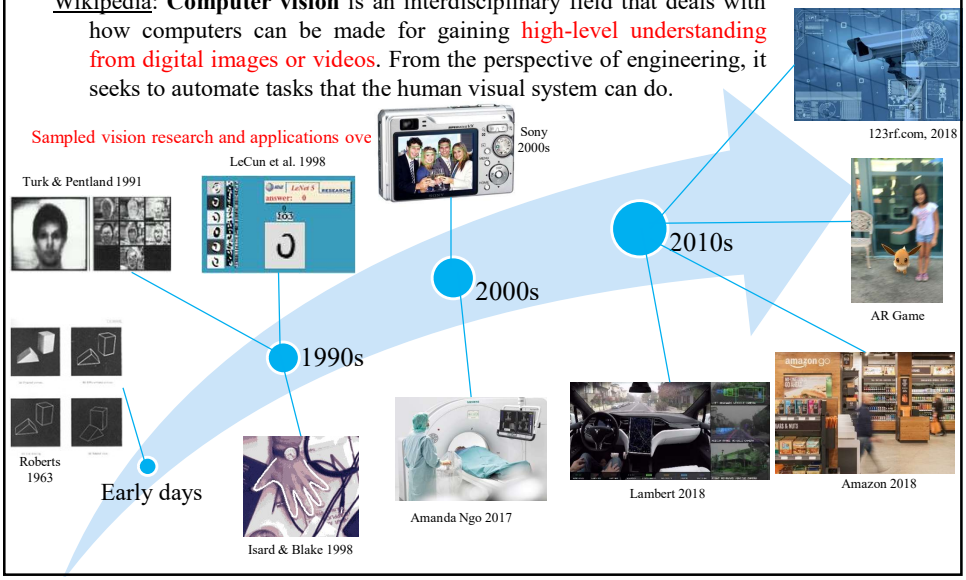
Haibin Ling

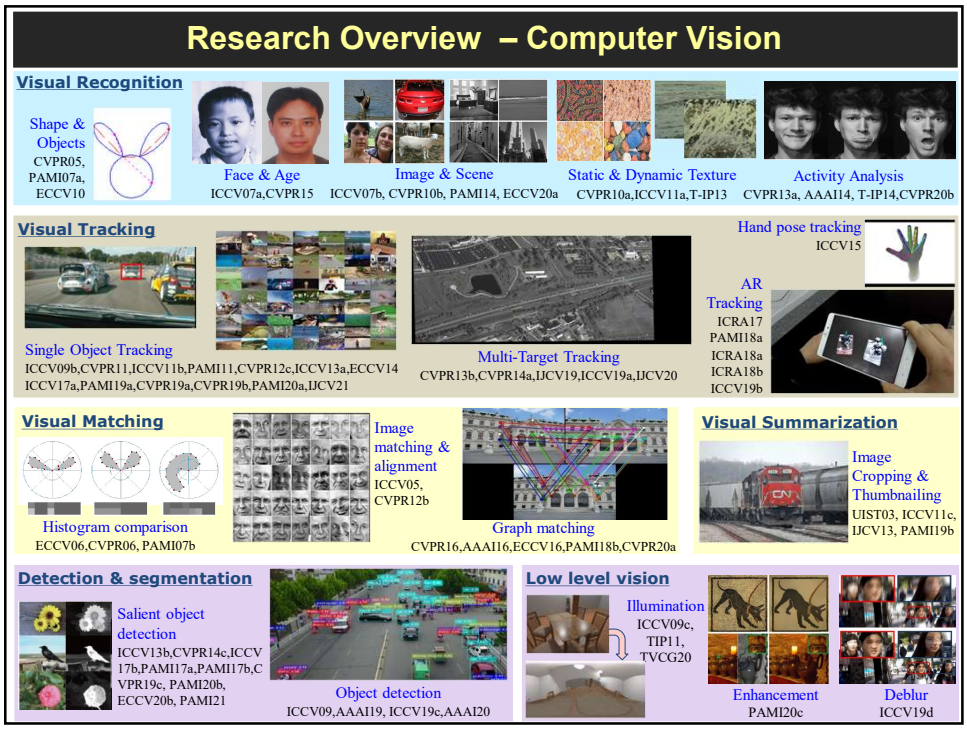
Stony Brook University

Feb. 17, 2021

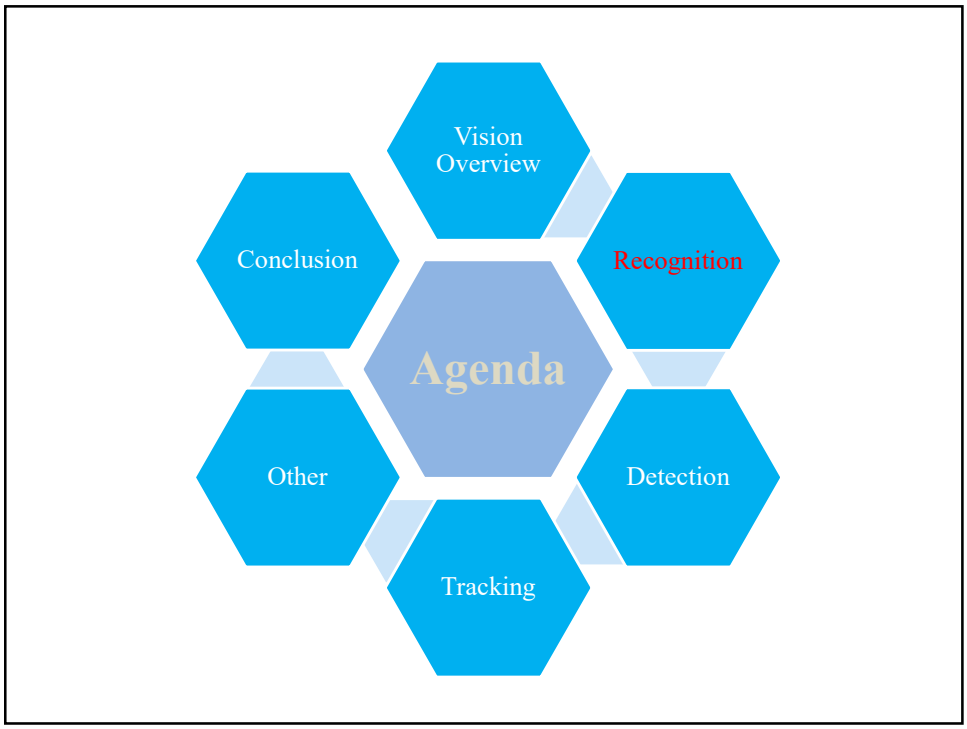
Computer Vision

Wikipedia: **Computer vision** is an interdisciplinary field that deals with how computers can be made for gaining high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do.



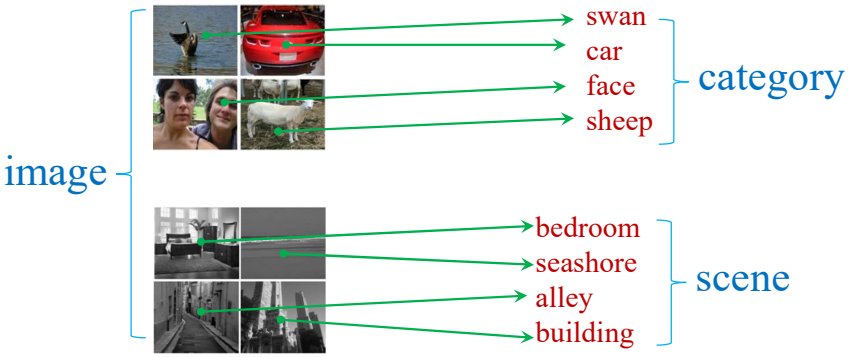



Research Overview – Applications
Biomedical image analysis
Breast image analysis
ISBI09,ISBI10, ISBI11,ISBI12
3D organ segmentation
CVPR08, CVPR09
Dental-based osteoporosis diagnosis
ISBI14,DMFR14,PR16,ISBI17 DMFR17,CHASE19,EMBC20
Neutrophils cell tracking
EMBC18
Curvilinear structure analysis
MICCAI09, CVPR14b, MICCAI16
Multi-modality fusion
PAMI21
Intelligent transportation
Nighttime traffic surveillance
T-ITS15,T-ITS18a
Pedestrian detection
T-ITS19a,T-ITS19b
Road crack detection
T-ITS'20
Traffic scene understanding
T-ITS18b,T-ITS18c,WACV19b
Visual privacy
Covert photo classification
CVPR12c,T-IP15,MVA17
De-identification
ICDAR11, IJCB14, JCST19
Retrieval
Re-identification
AAAI18, PR18, T-IP19
Cross-modal retrieval
PAMI20d
Species recognition
TAXON06, ECCV08, T-IP12
New Ongoing Research
SLAM & 3D reconstruction
T-RO19, ICRA20
Pose estimation
MICCAI 20
Projector compensation
Smart projection system
CVPR19d, ICCV19e, PAMI21
Application in chemical and material science


Vision Overview
Recognition
Detection
Tracking
Other
Conclusion
Agenda

Visual Recognition

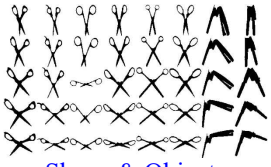
Input: visual data → Output: label(s)



Category & Scene Recognition
ICCV07b, CVPR10b, PAMI14, ECCV20a

Visual Recognition

Input: visual data → Output: label(s)



Shape & Objects
CVPR05, PAMI07a, ECCV10

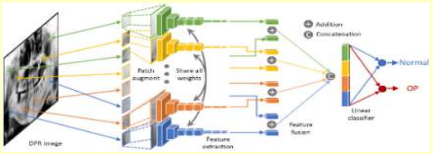
Face & Age
ICCV07a,CVPR15

Static & Dynamic Texture
CVPR10a,ICCV11a,T-IP13

Action & Interaction
CVPR13a, AAAI14, T-IP14,CVPR20b

Application – Visual Recognition

Medical Image



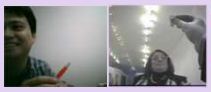
Dental-based osteoporosis prescreening
ISBI14,DMFR14,PR16,ISBI17,DMFR17,CHASE19,EMBC20

Retrieval

Cross-modal retrieval
PAMI20d

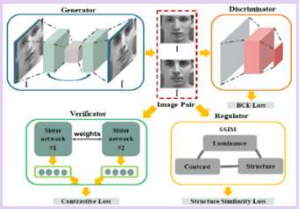
Species recognition
TAXON06, ECCV08, T-IP12

Visual privacy



Covert photo classification
CVPR12c,T-IP15,MVA17

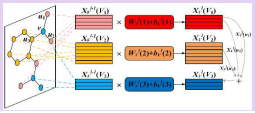
De-identification
ICDAR11, IJCB14, JCST19



Surveillance

Re-identification
AAAI18,PR18,T-IP19

Material science



Molecular property inference

Classical "Traditional" Methods

- Non-deep learning

Input → feature → classification

- Eigenface (Turk & Pentland, 1991)
- Haar (Viola & Jones, 2001)
- Shape Context (Belongie et al. 2002)
- SIFT (Lowe 2004)
- HOG (Dalal & Triggs, 2005)
- …

- Nearest neighbor
- Support vector machine
- AdaBoost
- Random forest
- Neural network
- …

Selected/learned/optimized Independently

AlexNet 2012

Winner of the ILSVRC2012 Challenge

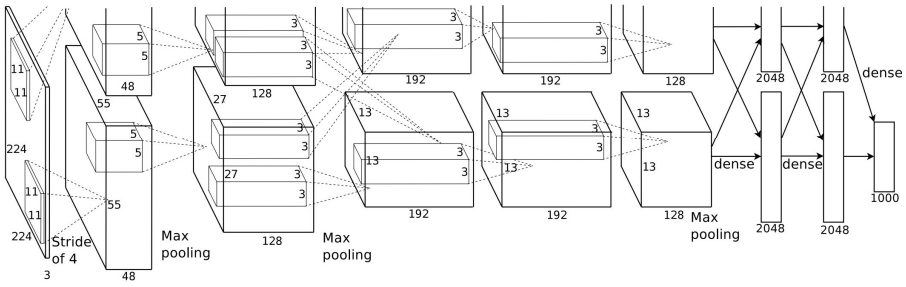


Figure 2: **Breakthrough:** 15.3% Top-5 error on ILSVRC2012, ~10% better than the next best (25.7%)

between th ion of responsibilities
at the bott her runs the layer-parts
528-dimensional, and
the number of neurons in the network's remaining layers is given by 253,440–186,624–64,896–64,896–43,264–4096–4096–1000.

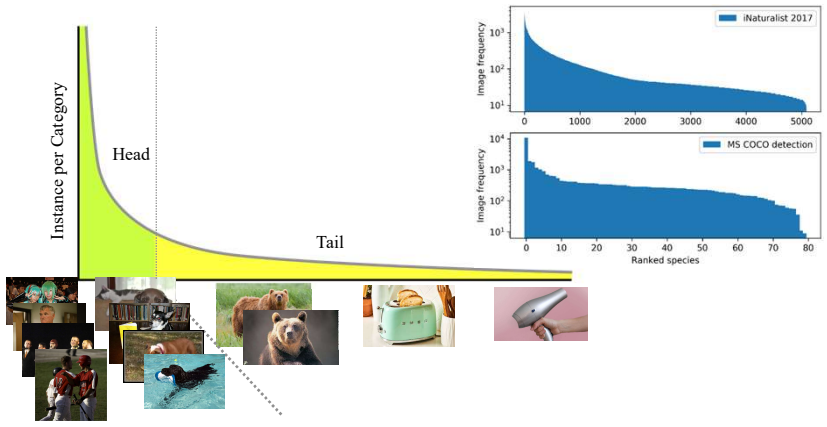
[Krizhevsky, Sutskever, Hinton. NIPS 2012]

Game Changing

- End-to-end
  - Joint and effective optimization
  - Flexibility
- Generalization to other computer vision tasks
  - Visual recognition is fundamental
  - Pretrain on ImageNet + fine tune for specific tasks
  - Extension to other vision tasks
    - Semantic understanding: detection, semantic segmentation, tracking, etc.
    - Geometric understanding: 3D reconstruction, simultaneous localization and mapping (SLAM), etc.
    - Low level vision: image enhancement, super resolution, etc.

Image Recognition for Long Tailed Data

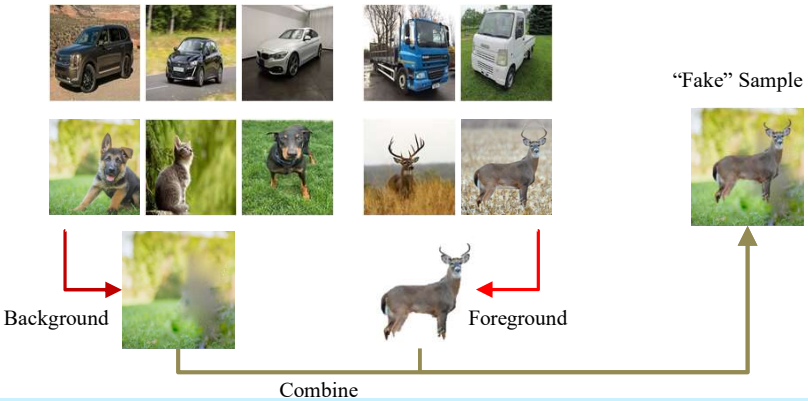
Long-tailed data is common in practice:



Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Data Augmentation – Image Space

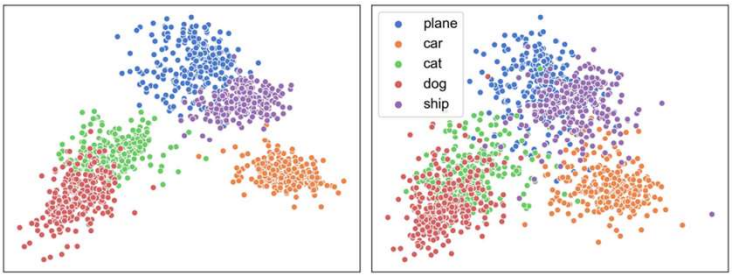
- Augment long-tailed class with synthetic data
- Combining tail class foreground and head class background
  - Artifacts and bias



Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Data Augmentation – Feature Space

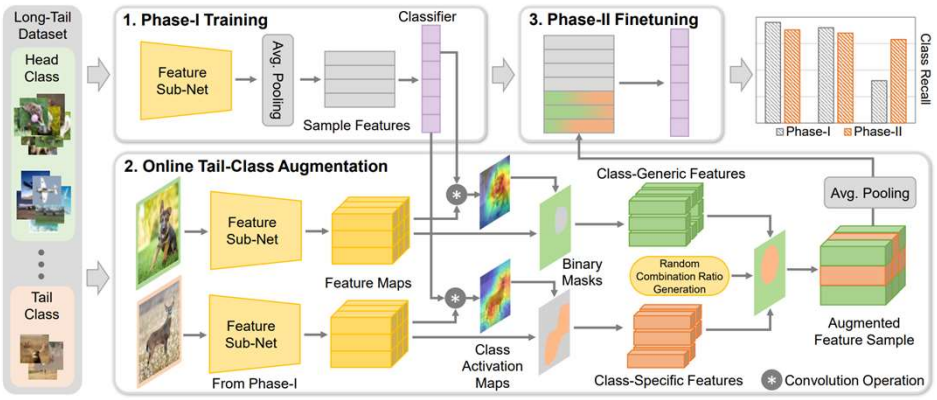
- Synthesize data in the feature space
- Class-specific feature vs. class-generic feature



Class-specific feature Class-generic feature

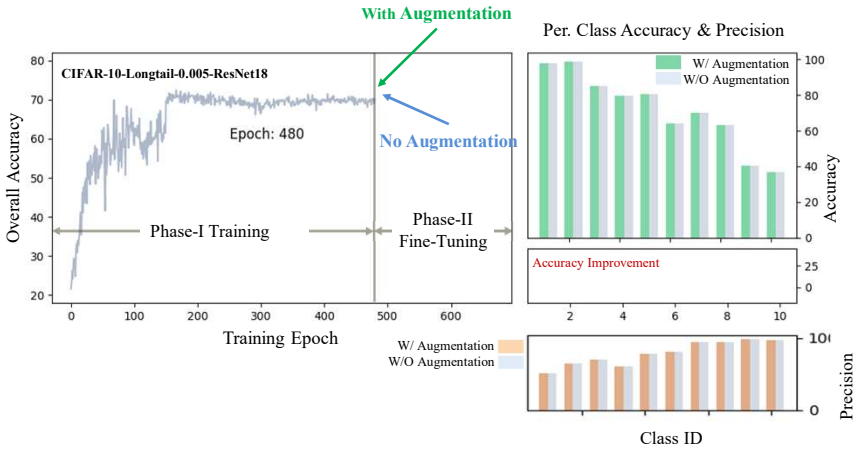
Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Algorithm Pipeline



Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Experiment: Learning Curve



Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Experiments: Small Dataset

Long-tailed CIFAR-10

	ResNet-18					ResNet-34				
IM	10	20	50	100	200	10	20	50	100	200
Baseline	90.73	87.24	82.32	75.16	70.22	91.03	87.32	82.74	78.58	71.42
CB [1] r = 0.9	90.79	86.61	81.9	75.16	69.16	91.03	87.18	82.48	75.99	70.0
CB [1] r = 0.999	90.54	86.83	81.81	76.4	69.83	90.74	87.24	81.66	74.85	70.08
CB [1] r = 0.9999	89.61	86.05	80.4	75.04	69.21	90.69	86.9	81.06	75.74	68.79
FL [2] r = 0.5	90.66	86.61	81.55	74.99	69.06	90.76	87.18	81.91	76.5	69.87
FL [2] r = 1.0	90.59	86.83	81.79	74.07	68.23	90.7	87.24	81.34	76.44	70.02
FL [2] r = 2.0	90.5	86.05	81.25	75.13	68.27	90.08	86.9	82.44	75.58	69.87
SLA [3]	-	-	-	-	-	89.58	-	-	80.24	-
Ours	**91.75**	**88.54**	**84.51**	**80.57**	**77.06**	**91.2**	**89.26**	**84.49**	**82.06**	**75.52**

Long-tailed CIFAR-100

	ResNet-18					ResNet-34				
IM	10	20	50	100	200	10	20	50	100	200
Baseline	62.59	57.09	48.55	43.65	38.87	63.87	57.55	48.07	43.55	37.5
CB [1] r = 0.9	63.1	57.02	48.15	43.51	38.58	64.14	58.03	48.44	42.94	38.84
CB [1] r = 0.999	61.76	55.3	44.28	32.19	26.61	63.05	54.13	40.89	32.65	26.2
CB [1] r = 0.9999	60.71	53.93	42.02	31.32	25.91	62.28	53.64	40.03	29.82	26.63
FL [2] r = 0.5	62.64	57.02	47.9	42.82	38.73	64.36	58.45	48.31	42.72	36.18
FL [2] r = 1.0	62.85	57.22	47.76	42.81	40.47	64.83	58.78	48.24	42.64	37.29
FL [2] r = 2.0	63.37	57.15	47.0	42.18	40.31	64.48	58.55	47.47	43.33	38.11
SLA [3]	-	-	-	-	-	59.89	-	-	45.53	-
Ours	**65.08**	**58.69**	**51.9**	**46.57**	**42.84**	**65.29**	**59.75**	**52.17**	**48.51**	**41.46**

$IMFactor = \frac{max(\{N_i\})}{min(\{N_i\})}$, N_i is the number of samples in $i-$ th class

Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

Experiments: Large Dataset

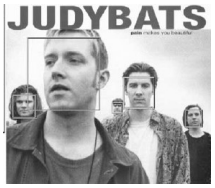
	ImageNet-LT				Places-LT			
	> 100 Many	≥100 & > 20 Medium	< 20 Few	Overall	> 100 Many	≥100 & > 20 Medium	< 20 Few	Overall
Plain Model	40.9	10.7	0.4	20.9	**45.9**	22.4	0.36	27.2
Lifted Loss	35.8	30.4	17.9	30.8	41.1	35.4	24	35.2
FL	36.4	29.9	16	30.5	41.1	34.8	22.4	34.6
Range Loss	35.8	30.3	17.6	30.7	41.1	35.4	23.2	35.1
FSLwF	40.9	22.1	15	28.4	43.9	29.9	**29.5**	34.9
OLTR	43.2	**35.1**	**18.5**	**35.6**	44.7	37	25.3	35.9
Ours	**47.3**	31.6	14.7	35.2	42.8	**37.5**	22.7	**36.4**
Ours+FL	47	31.3	16.8	35.3	42.2	36.4	24	36

		iNaturalist 2017	iNaturalist 2018
Baseline	ResNet-50	60.5	62.27
	ResNet-101	61.81	65.19
	ResNet-152	65.12	66.17
CB	ResNet-50	58.08	61.12
	ResNet-101	60.94	63.88
	ResNet-152	64.75	66.97
Our	ResNet-50	**61.96**	**65.91**
	ResNet-101	**64.16**	**68.39**
	ResNet-152	**66.58**	**69.08**

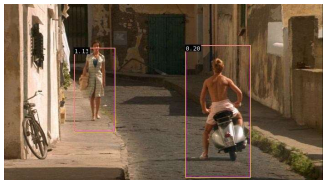
Feature Space Augmentation for Long Tailed Data. Chu, Bian, Liu, & Ling, ECCV 2020

What is Visual Detection?

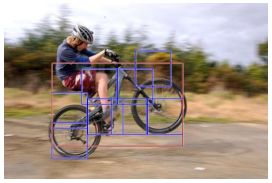
- Locate object(s) in an input image



Viola & Jones IJCV 2004



Dalal & Triggs CVPR 2005

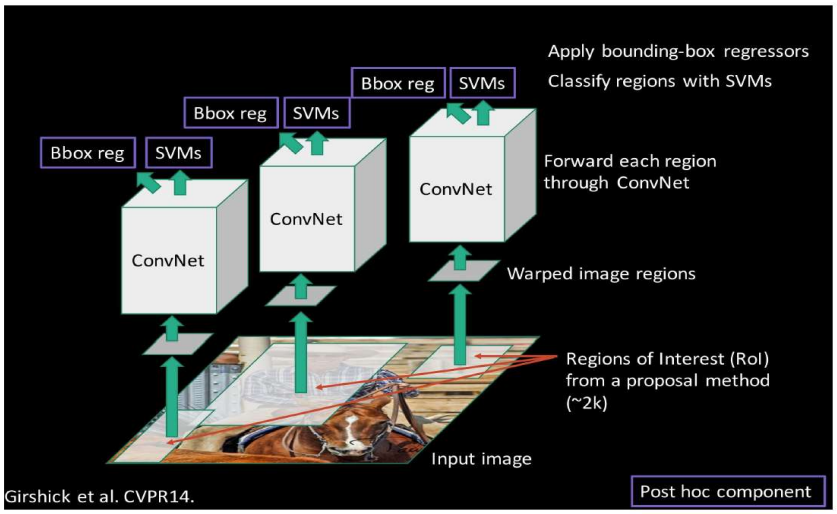


Felzenszwalb, Girshick, McAllester & Ramanan, PAMI 2010

- Extensions
  - Object segmentation
  - Object detection in videos
  - Salient object detection

Yang, Fan, Chu, Blasch & Ling

Deep Learning Solutions

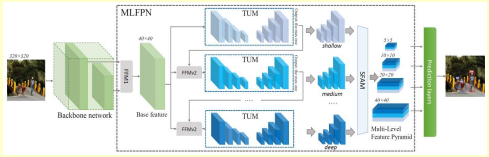


R-CNN after region proposal

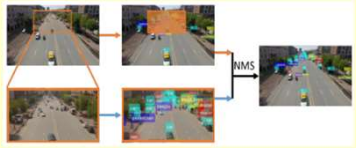
Slide credit: Ross Girschick

Our Studies on Visual Detection

Model-based Object Detection



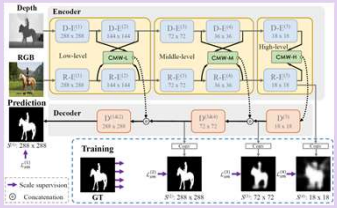
Generic object detection
ICCV09, AAAI19, AAAI20



Specific application scenario
ICCV19c

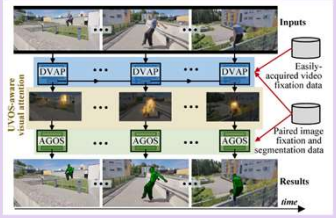
Salient Object Detection

Detection from Color Image
ICCV13b,CVPR14c,ICCV 17b,PAMI17a, PAMI17b



Detection from RGB-D Input
ECCV20b

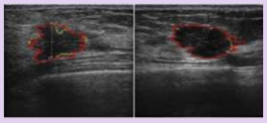
Video Object Detection



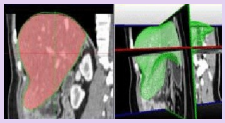
CVPR19c, PAMI21b

Application of Visual Detection

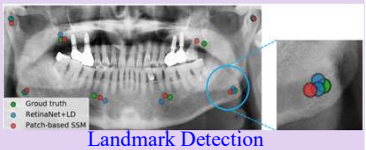
Medical Image



Breast image analysis
ISBI09,ISBI10, ISBI11,ISBI12

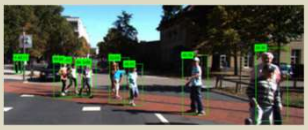


3D organ segmentation
CVPR08, CVPR09

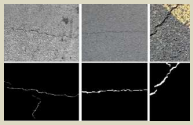


Landmark Detection
EMBC20

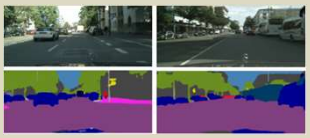
Deriving Assistance



Pedestrian detection
T-ITS19a,T-ITS19b



Road crack detection
T-ITS19c



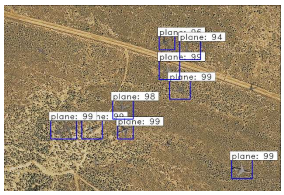
Traffic scene understanding
T-ITS18b,T-ITS18c,WACV19b,T-ITS21

Object Detection in Aerial Images

- Object detection in aerial image
  - Detection object in an image taken by drones, or general UAVs, equipped with cameras.
  - A wide range of applications, including agricultural, aerial photography, fast delivery, and surveillance.

UAVDT

VisDrone

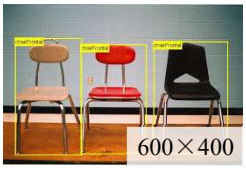


DOTA

Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Issues of Existing Detectors

- State-of-the-art detectors on aerial images
  - Faster RCNN+FPN (ResNet50)
  - **36.7** AP on COCO
  - **21.4** AP on VisDrone



- Issues of aerial images
  - Large scale image
  - Many objects
  - Small object
  - Large scale change within an image

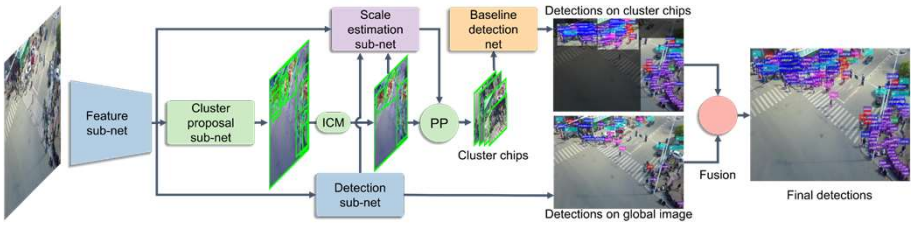
Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Motivation

- Clustered detection
  - Objects are unevenly distributed and locally clustered → clustering before detection
  - Within cluster scale variation is low → cluster dependent scale normalization

Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Method (ClusDet)



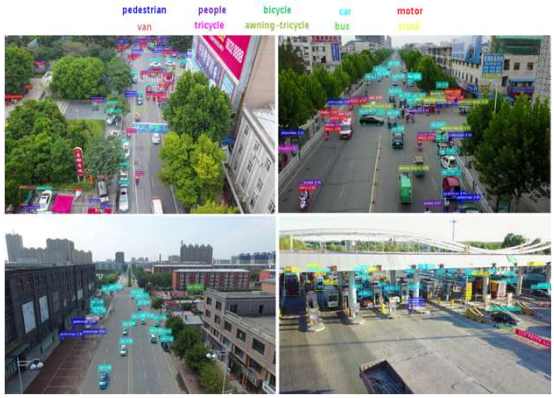
ICM: iterative cluster merging
PP: partition and padding

Clustered object detection (ClusDet) network

- Three key components:
  - Cluster proposal subnet (CPNet)
  - Scale estimation subnet (ScaleNet)
  - Dedicated detection network (DetecNet)
- Final detection results:
  - Fusing detections from cluster chips and global image.

Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Experiment on VisDrone



Method	backbone	AP	AP_{50}
RetinaNet+FPN	RNet50	13.9	23.0
RetinaNet+FPN	RNet101	14.1	23.4
RetinaNet+FPN	RNeXt101	14.4	24.1
FRCNN+FPN	RNet50	21.4	40.7
FRCNN+FPN	RNet101	21.4	40.7
FRCNN+FPN	RNeXt101	21.8	41.8
FRCNN+FPN *	RNeXt101	28.7	51.8
FRCNN+FPN+EIP	RNet50	21.1	44.0
FRCNN+FPN+EIP	RNet101	23.5	46.1
FRCNN+FPN+EIP	RNeXt101	24.4	47.8
FRCNN+FPN +EIP*	RNeXt101	25.7	48.4
ClusDet	RNet50	26.7	50.6
ClusDet	RNet101	26.7	50.4
ClusDet	RNeXt101	28.4	53.2
ClusDet*	RNeXt101	**32.4**	**56.2**

* Multi-scale inference & bounding box voting utilized in testing

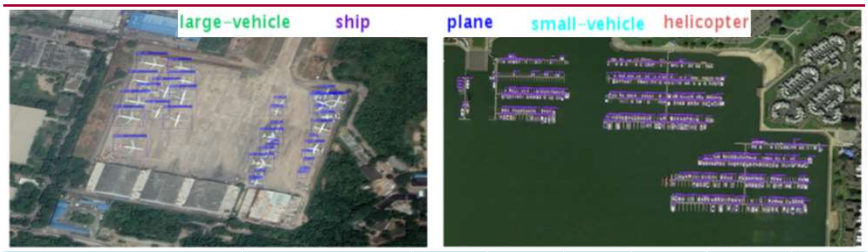
Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Experiment on UAVDT

Method	backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
R-FCN	ResNet50	7.0	17.5	3.9	4.4	14.7	12.1
SSD	N/A	9.3	21.4	6.7	7.1	17.1	12.0
RON	N/A	5.0	15.9	1.7	2.9	12.7	11.2
FRCNN	VGG	5.8	17.4	2.5	3.8	12.3	9.4
FRCNN+FPN	ResNet50	11.0	23.4	8.4	8.1	20.2	26.5
FRCNN+FPN+EIP	ResNet50	6.6	16.8	3.4	5.2	13.0	17.2
ClusDet	ResNet50	**13.7**	**26.5**	**12.5**	**9.1**	**25.1**	**31.2**

Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

Experiment on DOTA



Method	Backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
RetinaNet+FPN+EIP	ResNet50	24.9	41.5	27.4	9.9	32.7	30.1
RetinaNet+FPN+EIP	ResNet101	27.1	44.4	30.1	10.6	34.8	33.7
RetinaNet+FPN+EIP	ResNeXt101	27.4	44.7	29.8	10.5	35.8	32.8
FRCNN+FPN+EIP	ResNet50	31.0	**50.7**	32.9	16.2	37.9	37.2
FRCNN+FPN+EIP	ResNet101	31.5	50.4	36.6	16.0	**38.5**	38.1
ClusDet	ResNet50	**32.2**	47.6	**39.2**	16.6	32.0	**50.0**
ClusDet	ResNet101	31.6	47.8	38.2	15.9	31.7	49.3
ClusDet	ResNeXt101	31.4	47.1	37.4	**17.3**	32.0	45.4

Clustered Object Detection in Aerial Images, Yang, Fan, Chu, Blasch, & Ling, ICCV 2019

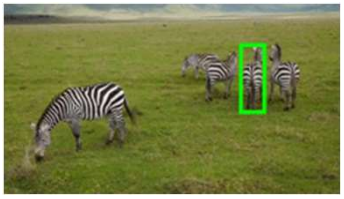
What is Visual Tracking

Visual Tracking

Keep track of something across time

t

Object defined manually or by detection

Model-free single object tracking

ICCV 2009, CVPR 2011, PAMI 2011, ICCV 2011, CVPR 2012, ICCV 2013, ECCV 2014, ICCV 2017, PAMI 2019, CVPR 2019a, CVPR2019b, PAMI2020
with C. Bao, E. Blasch, H. Fan, J. Gao, W. Hu, H. Ji, X. Mei, Y. Wu, J. Xing, F. Yang, et al.

What is Visual Tracking

Visual Tracking

Keep track of something across time

t

Plane/Pose

ICRA 2017, PAMI 2018, ICRA 2018a, ICRA 2018b
with L. Chen, P. Liang, T. Wang, et al.

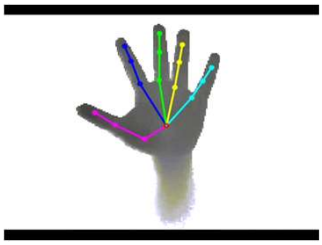
What is Visual Tracking

Visual Tracking

Keep track of something across time

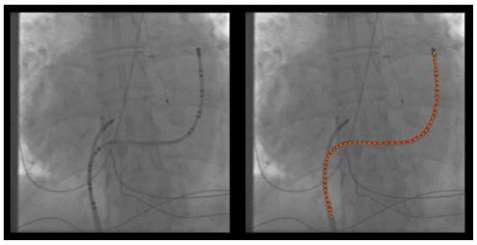
Deformable structures

Hand pose



ICCV 2015
with P. Li, X. Li, C. Liao

Curvilinear structure



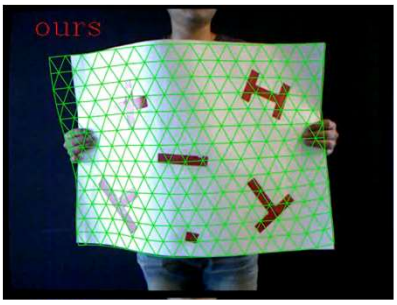
CVPR 2014, MICCAI 2016
with E. Cheng, P. Chu, Y. Pang, Y. Zhu

What is Visual Tracking

Visual Tracking

Keep track of something across time

Deformable surface



With Wang, Lang, Feng, & Hou, ICCV 2019

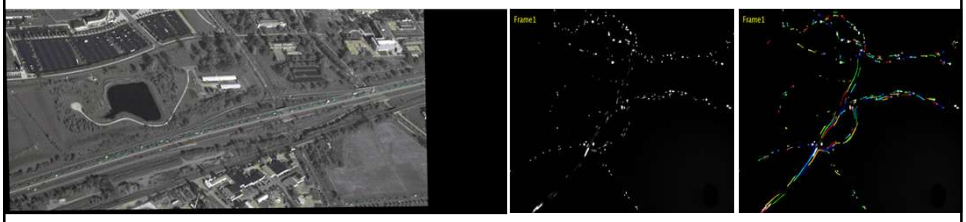
What is Visual Tracking

Visual Tracking

Keep track of something across time

t

Multiple targets



Highway vehicles

Neutrophil cell motion

CVPR 2013, CVPR 2014, EMBC 2018, IJCV 2019
with P. Chu, W. Hu, M.F. Kiani, Y. Pang, X. Shi, F. Soroush, J. Xing, *et al.*

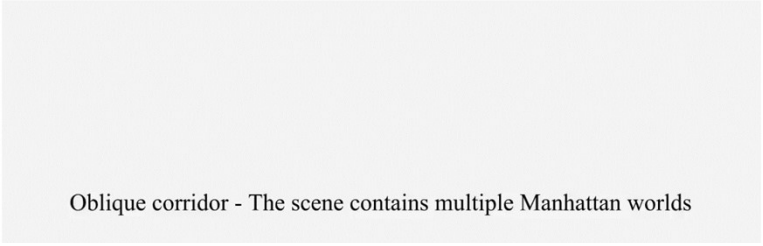
What is Visual Tracking

Visual Tracking

Keep track of something across time

t

Camera pose & environment geometry (SLAM)

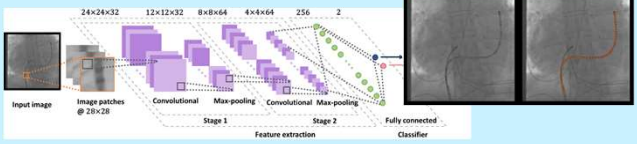


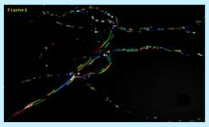
Oblique corridor - The scene contains multiple Manhattan worlds

T-RO 2019
with D. Zou, Y. Wu, L. Pei, W. Yu

Application of Visual Tracking

Biomedical image analysis





Curvilinear structure analysis MICCAI09,CVPR14b,MICCAI16

Neutrophils cell tracking
EMBC18

Traffic Surveillance

Highway vehicle monitoring
CVPR13b, CVPR14a, IJCV19, IJCV20

Nighttime traffic surveillance
T-ITS15,T-ITS18a

Augmented Reality

AR Game
ICRA17, PAMI18a, ICRA18a, ICRA18b, ICCV19b

3D Cylinder Tracking
T-CSVT20

Benchmark

LaSOT: Large Scale Single Object Tracking Benchmark

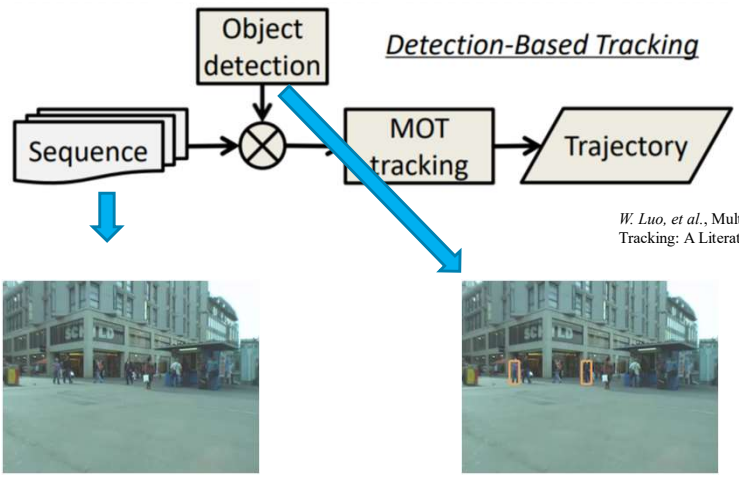
Fan, Bai, Lin, Yang, Chu, Deng, Yu, Harshit, M. Huang, Xu, Liao, Yuan, & Ling

CVPR 2019 | IJCV 2021

- Large: 1,550 seq, 3.52M+ frames
- Long-term: >2,000 frames per seq
- Diversity: ~85 categories
- Balance: ~20 seq/category
- Quality: per frame manually annotation
- Language: text description per seq

LaSOT

Multiple Object Tracking

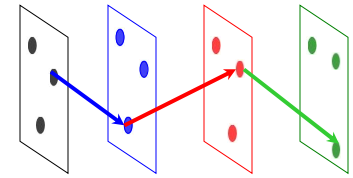


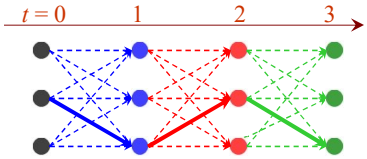
W. Luo, et al., Multiple Object Tracking: A Literature Review

FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking, Chu & Ling, ICCV 2019

Problem Formulation

A toy example: K= 3, N=3





Input:

- K+1 frames
  - "+1" for convenience
- Each with N detections
  - Usually with false alarms
  - N can be frame-dependent

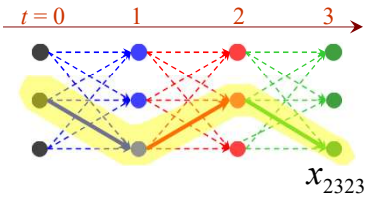
Output:

- Trajectories over time

FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking, Chu & Ling, ICCV 2019

Multi-Dimensional Assignment

(K+1)-dimensional assignment,
(K+1)-partite problem



$$\min_{X=\{x_{i_0 i_1 \ldots i_K}\}} \sum_{i_0}^{N}\sum_{i_1}^{N}\ldots\sum_{i_K}^{N} c_{i_0 i_1 \ldots i_K} x_{i_0 i_1 \ldots i_K}$$

Trajectory indicator

(K+1)-order trajectory cost

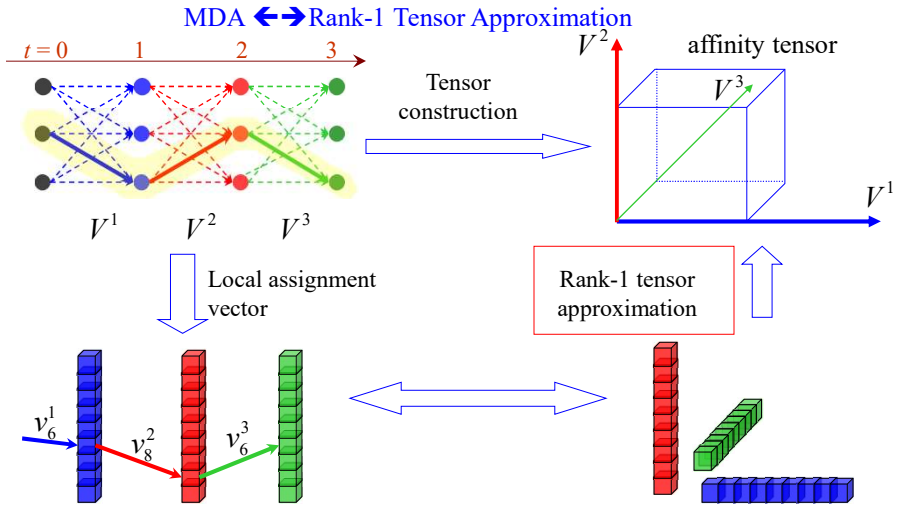
$$s.t. \begin{cases} \sum_{\{i_0, i_1, \ldots, i_K\}/i_k} x_{i_0 i_1 \ldots i_K} = 1, & k = 0,1,\ldots,K, \quad i_k = 1,\ldots,N \\ x_{i_0 i_1 \ldots i_K} \in \{0,1\}, & i_0,\ldots,i_K = 1,\ldots,N \end{cases}$$

Trajectory disjoint constraints

Binary constraints

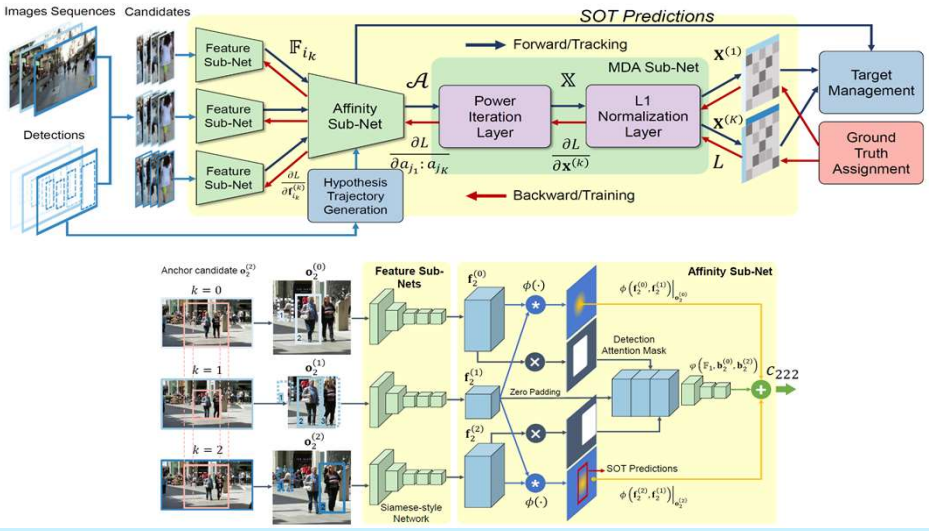
FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking, Chu & Ling, ICCV 2019

From MDA to Tensor Analysis



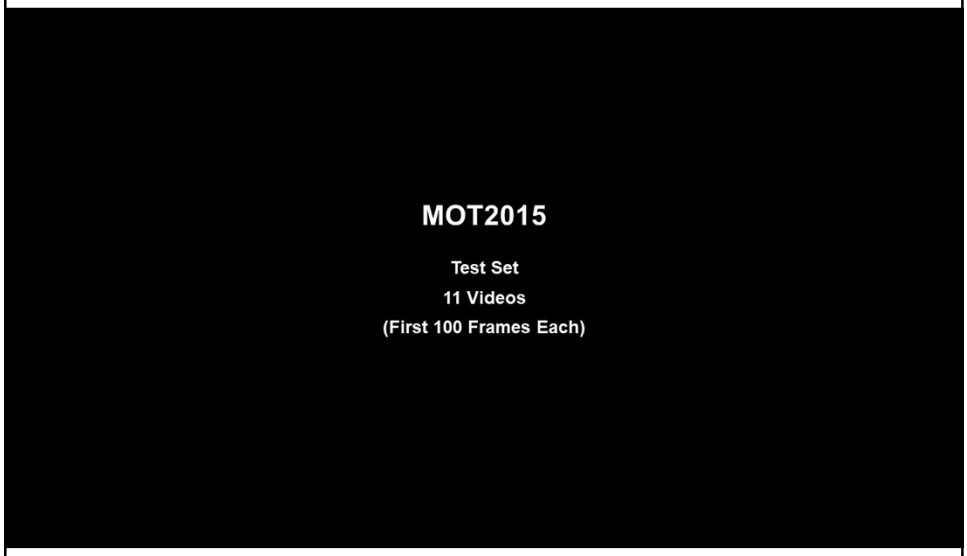
Shi, Ling, et al., CVPR 2013, CVPR 2014, CVPR 2016, IJCV 2019

End-to-End Multi-Object Tracking



FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking, Chu & Ling, ICCV 2019

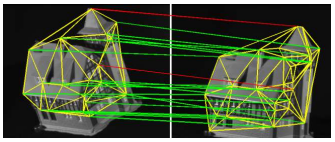
Qualitative Results

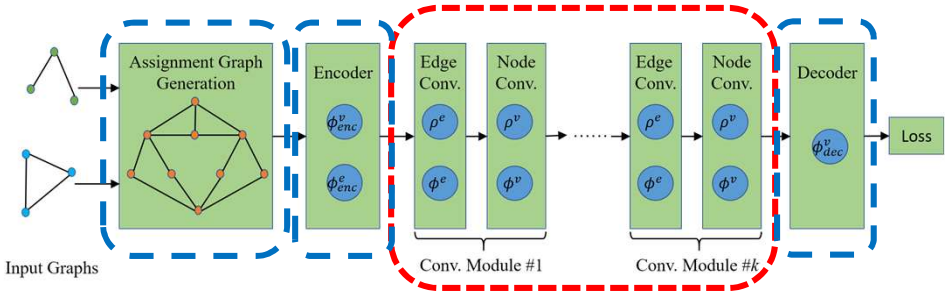


FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking, Chu & Ling, ICCV 2019

Graph Matching

- Graph matching:
  - matching vertices and edges.
- Graph neural network for graph matching.

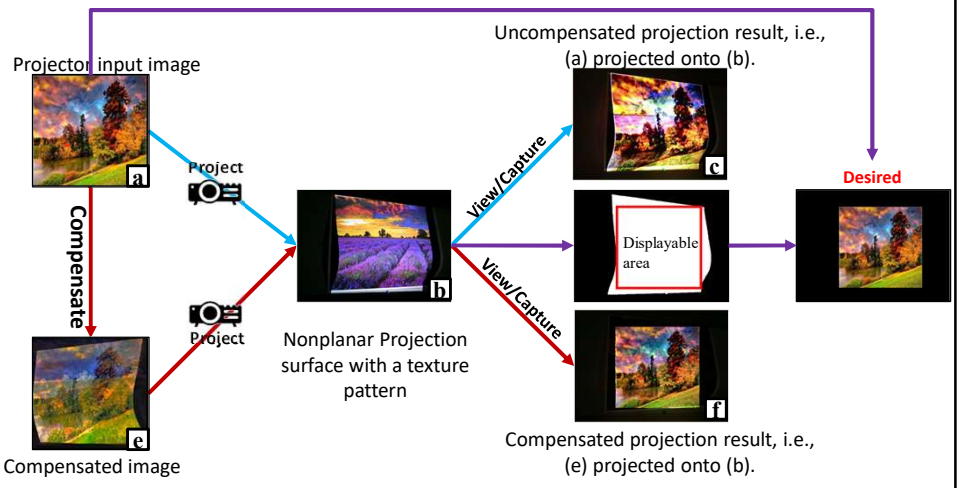




Graph Network（GN） Blocks

Learning Combinatorial Solver for Graph Matching. Wang, Liu, Li, Jin, Hou, & Ling, CVPR 2020

Projector Compensation (ProCams)



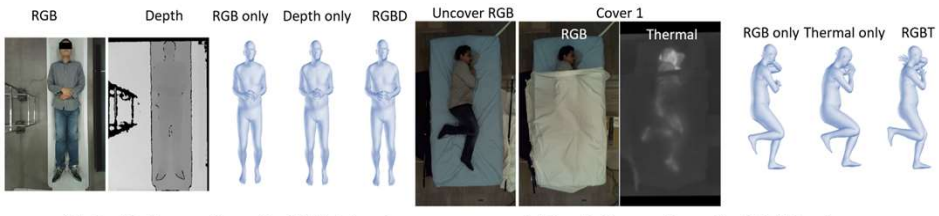
End-to-end Full Projector Compensation

Huang, Sun, & Ling, CVPR 2019, ICCV 2019, PAMI in press.

Huang, Sun, & Ling, CVPR 2019, ICCV 2019, PAMI in press.

3D Body Reconstruction

- Recovering patient body shape (mesh) from multi-modal inputs.



(a) Qualitative results on the SCAN dataset

(b) Qualitative results on the SLP dataset

Robust Multi-modal 3D Patient Body Modeling. Yang, Li, Georgakis, Karanam, Chen, Ling, & Wu, MICCAI 2020.

Conclusion

- We have summarized our work on various computer vision tasks and applications.
- End-to-end modeling has dominantly outperforms traditional strategies in computer vision tasks.
- Performances in many vision tasks have surged so as to largely boost their deployment in real-world applications.
- Challenges such as data imbalance and data insufficiency encourage integrating domain knowledge into network design, data augmentation, etc.

Acknowledgement

- Students/Postdocs:
  - Erkang Cheng, Peng Chu, Liang Du, Heng Fan, Bingyao Huang, Sarah Lehman, Tatyana Nuzhnaya, Xinchu Shi, Fariborz Soroush, Tao Wang, Wenguan Wang, Yi Wu, Fan Yang
- Collaborators:
  - Erik Blasch, Xiao Bian, Weiming Hu, Mohammad Kiani, Vasileios Megalooikonomou, Jianbing Shen, Ziyan Wu, Jie Yang
- Support:
  - National Science Foundation (NSF)
  - Yahoo Faculty Research and Engagement Program Award
  - Amazon Machine Learning Research Award