

Briefing on CVPR 2021

Futurewei Tech: Yaohui Ding

June 30, 2021

CVPR 2021 – June 19-25

❑ # of Papers: 1663/7015 (1470/6656 in 2020)

- Huawei presenter: > 19

❑ AI/CV Trend in recent years:

- From learning representations to **high-level tasks**
- more on 3D construction, pose/motion/action, GAN, auto-drive, etc.

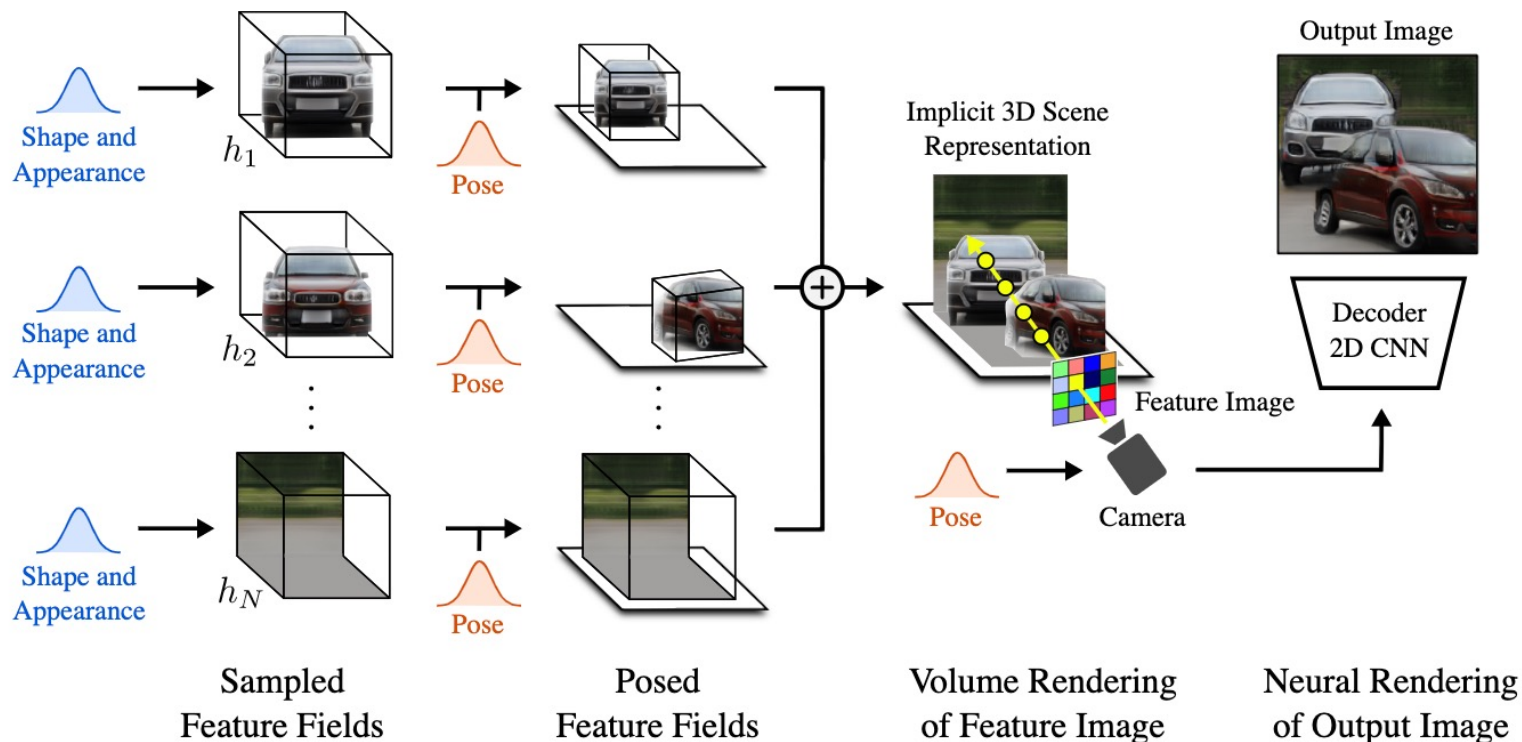
❑ My briefing:

- From the **awarded** (1 best, 2 honorable mentions, 32 candidates)
- A new task: **open-set** panoptic segmentation (OPS)
- Transformer in Vision

Best Paper: GIRAFFE

“GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields”

<https://m-niemeyer.github.io/project-pages/giraffe/index.html>



Takeaways:

- Controllable generative images
- input 2D, sample 3D info, render 2D
- Random sampling to generate feature field, but fix the pose at the rendering
- Single GPU training

Best Paper Honorable Mentions 1/2

1 – FAIR: Exploring Simple Siamese Representation Learning

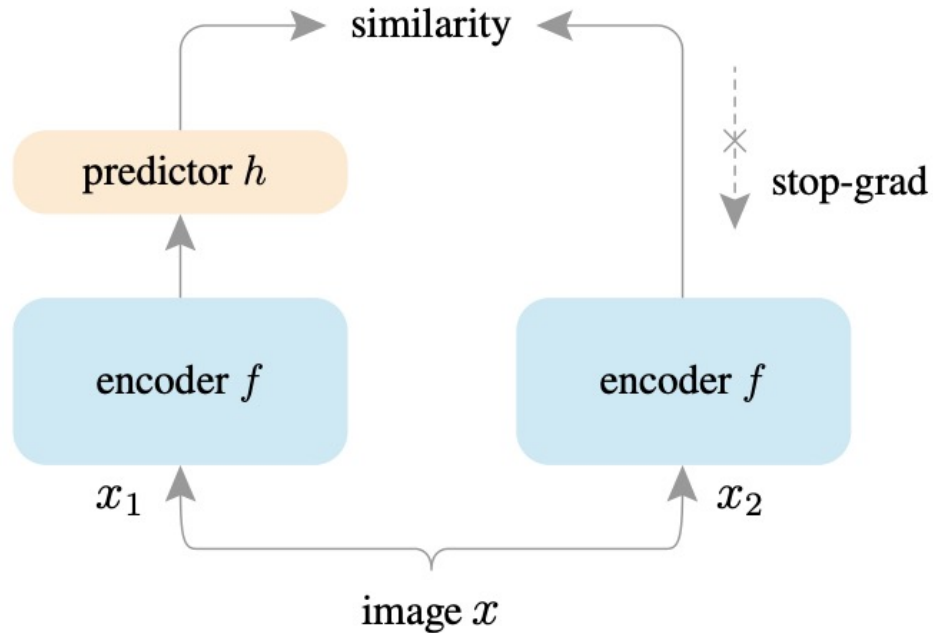


Figure 1. **SimSiam architecture.** Two augmented views of one image are processed by the same encoder network f (a backbone plus a projection MLP). Then a prediction MLP h is applied on one side, and a stop-gradient operation is applied on the other side. The model maximizes the similarity between both sides. It uses neither negative pairs nor a momentum encoder.

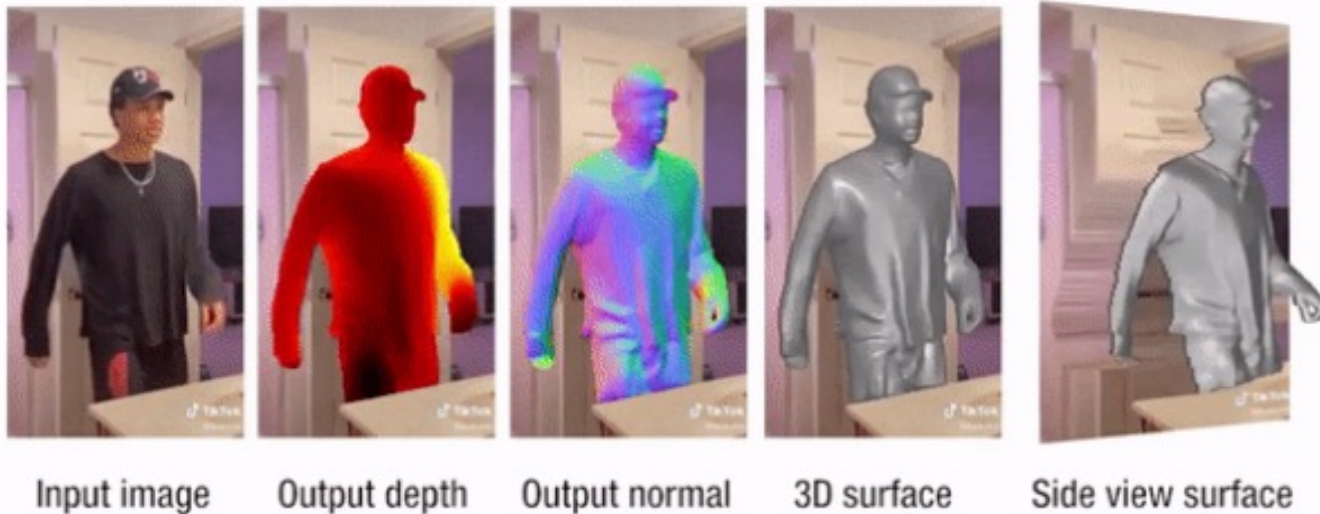
Takeaways:

- un/self-supervised learning
- Siamese Network
- Augmentation, but void “collapsing”
- what really works?
 - i) negative pairs?
 - ii) big batch?
 - iii) Momentum embedding
 - iv) All above are NOT necessary**

Best Paper Honorable Mentions 2/2

2 - “Learning High Fidelity Depths of Dressed Humans by Watching Social Media Dance Videos” by University of Minnesota

https://www.yasamin.page/hdnet_tiktok



Takeaways:

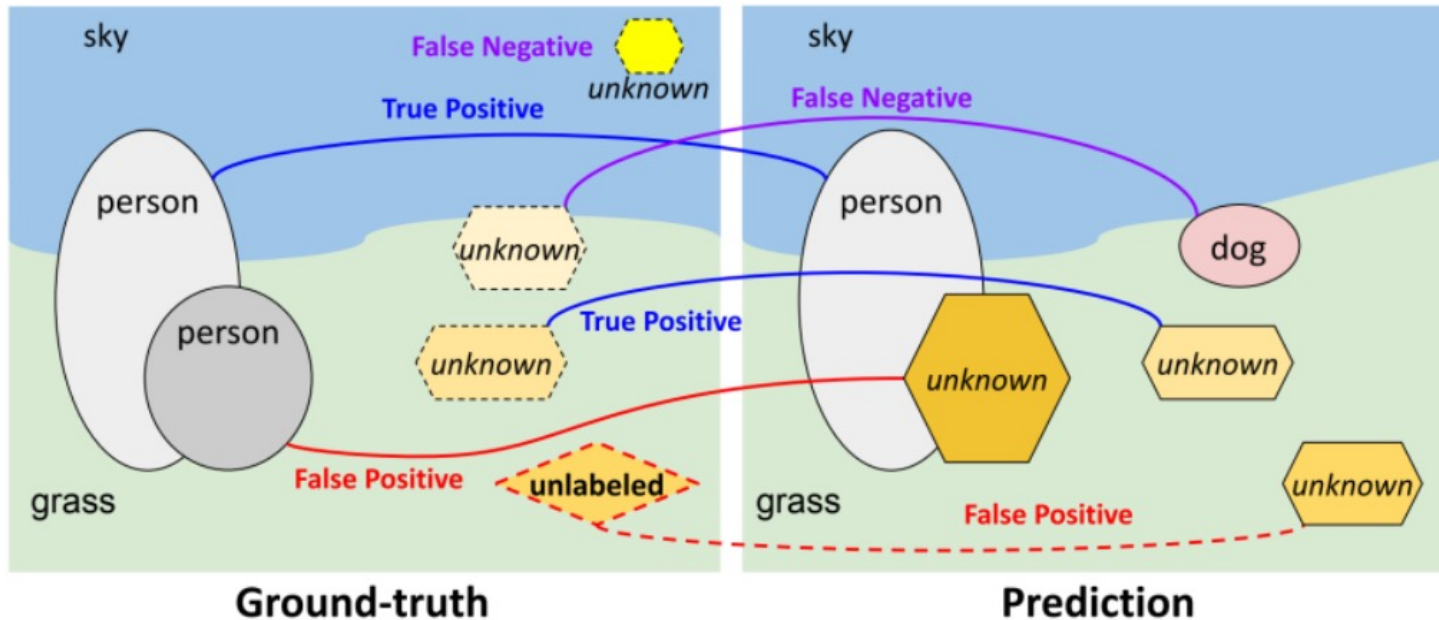
- Best results on depth prediction
- Limit ground truth until a new data source
- Training data are from a dance challenge in **TikTok**
- Rich but similar movement, but diverse appearance, clothing, background

Briefing

- ❑ # of Papers: 1663/7015
 - Huawei: >19
- ❑ AI/CV Trend: more on high-level tasks
 - 3D construction, pose/motion/action, GAN, auto-drive, etc.
- ❑ My briefing:
 - From the **awarded** (1 best, 2 honorable mentions, 32 candidates)
 - A new task: **open-set** panoptic segmentation (OPS)
 - **Transformer** in Vision

A New Task: Open Set Panoptic Segmentation

- “Exemplar-Based Open-Set Panoptic Segmentation Network”
by 首尔大学&Adobe



Takeaways:

- Panoptic segmentation overwhelms instance/semantic in academic
- Limited annotation leads to false positive on unlabeled pixels
- Open-set task is closer to realistic scenarios

- Related: Towards Open World Object Detection (oral paper)

- Identify objects that have not been introduced to it as ‘unknown’
- When the labels are received, incrementally learn these identified unknown categories **without forgetting** previously learned classes

Transformer in Vision

❑ > 43 papers in CVPR2021

❑ Pro:

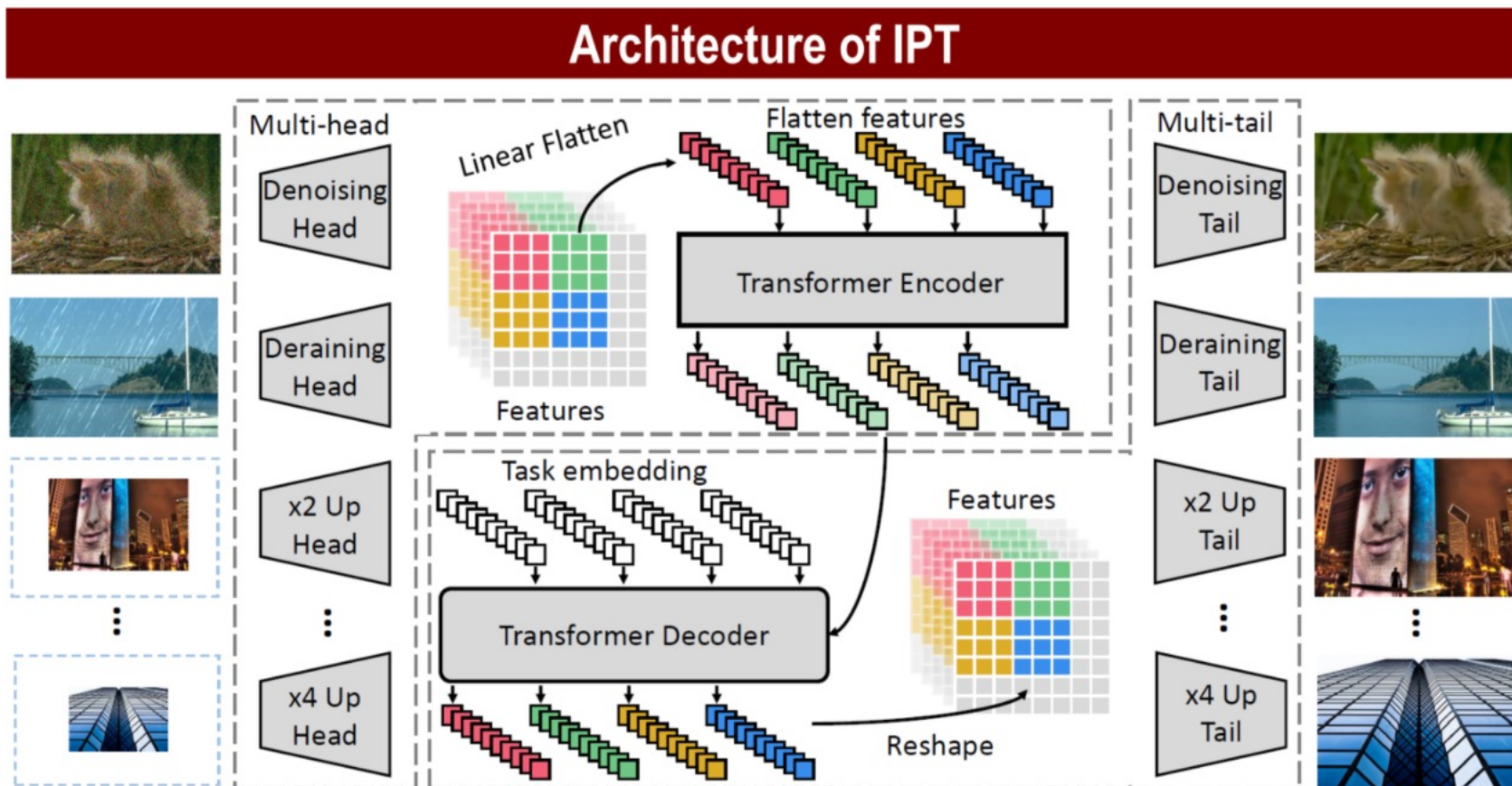
- Easily borrow tips from NLP
- Long range context compared to CNN
- Embedding-based Multi-Media Fusion
- Pretrain for multi-task

❑ Con:

- Deploy
- Lack of operation-based optimization in CV context
- Is performance really better than CNN?

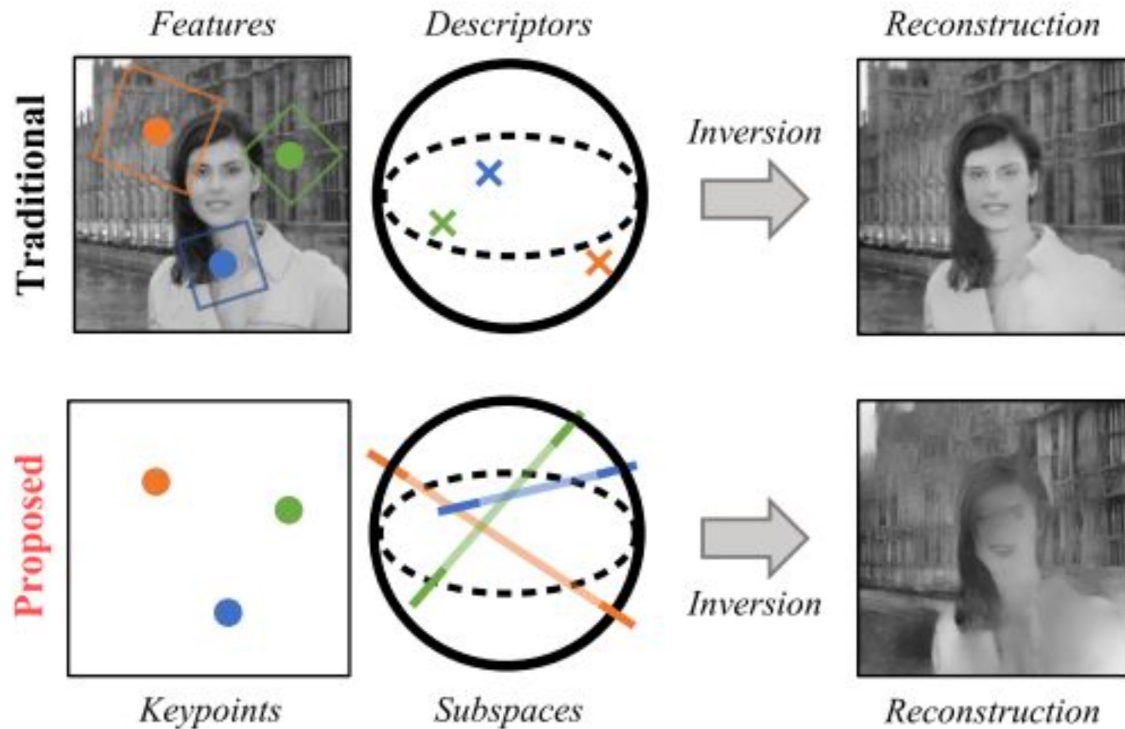
Transformer in Vision

“Pre-Trained Image Processing Transformer” by 北大, 悉尼大学, 华为



Among 32 Best Paper Candidates

- **Privacy-Preserving Image Features via Adversarial Affine Subspace Embeddings** by 苏黎世联邦理工学院&微软

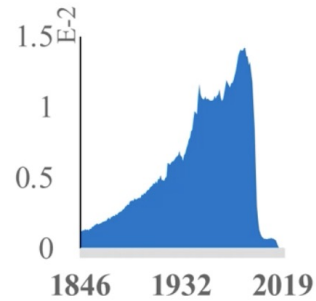


- Face matching, avoid leaking of feature space
- Lifting new features to subspace, then match

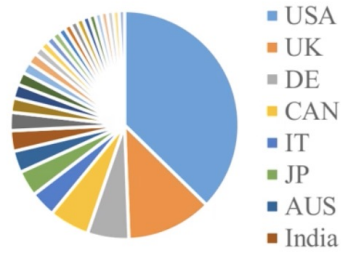
Related to Face Detection

- “WebFace260M: A Benchmark Unveiling the Power of Million-Scale Deep Face Recognition”

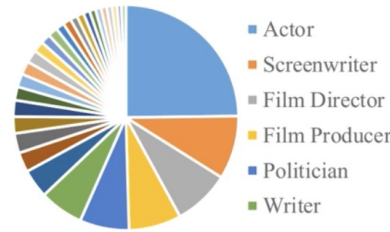
Training Set



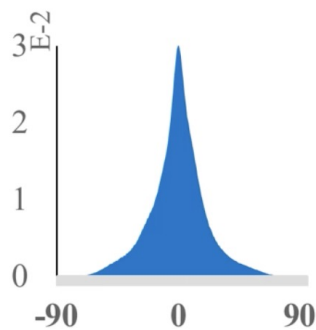
(a) Date of Birth



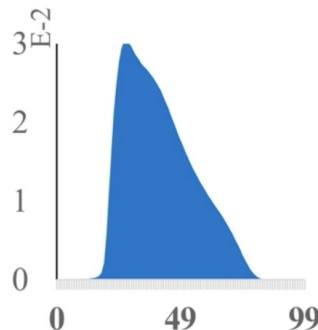
(b) Nationality



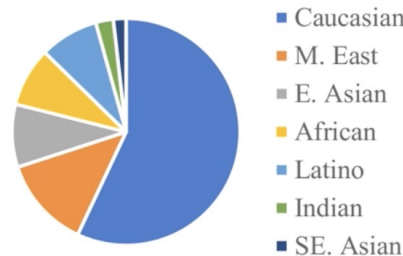
(c) Profession



(a) Pose



(b) Age



(c) Race

- First public face database > 10M
- 260M images from 4M sub
- cleaned 42M images from 2M sub
- Rich attributes
- lccv2021 Masked face detection

Thank You.

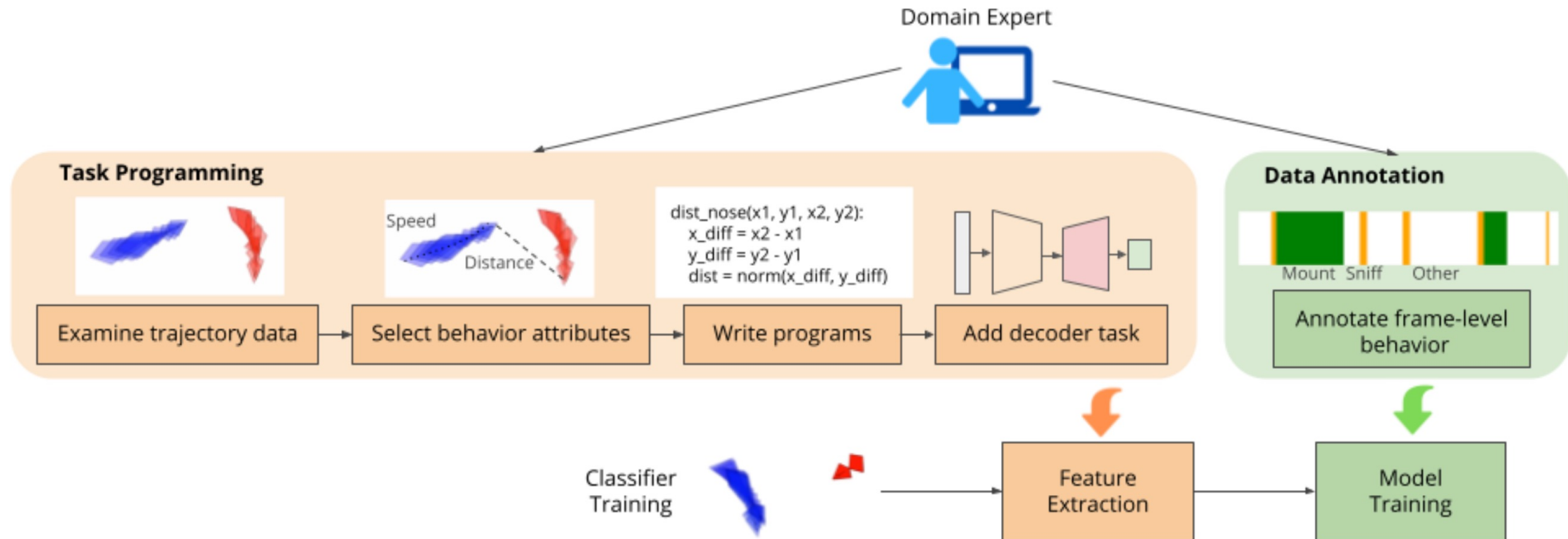
**Copyright © 2019 Futurewei Technologies, Inc.
All Rights Reserved.**

The information in this document may contain predictive statements including, without limitation, statements regarding the future financial and operating results, future product portfolio, new technology, etc. There are a number of factors that could cause actual results and developments to differ materially from those expressed or implied in the predictive statements. Therefore, such information is provided for reference purpose only and constitutes neither an offer nor an acceptance. Futurewei may change the information at any time without notice.



Supplementary 1: Best Student Paper

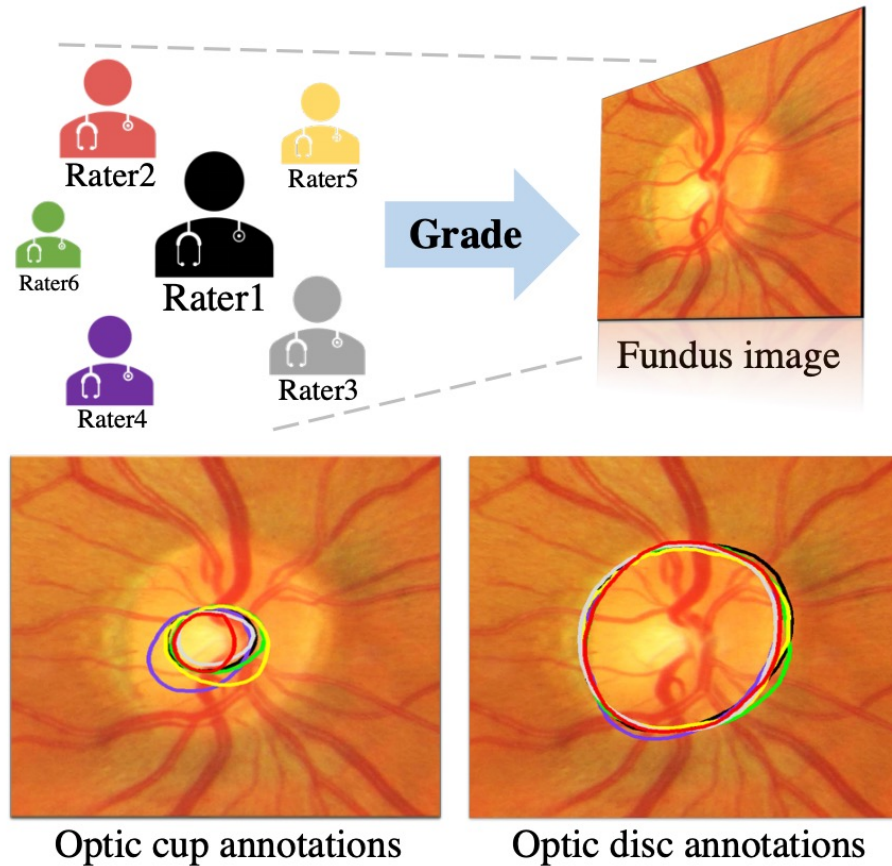
- “Task Programming: Learning Data Efficient Behavior Representations” by Caltech & Northwestern (美)



- Behavior analysis
- Goal is to reduce annotation effort
- Simulate the behavior for similar but simple task
- Annotate on fewer real behavior

Supplementary 2. Medical Image Segmentation

- “Learning Calibrated Medical Image Segmentation via Multi-Rater Agreement Modeling by 阿尔伯塔大学&腾讯天衍实验室&哈工大&阿姆斯特丹大学”



- Multi-Rater Agreement
- “Expertness” as prior knowledge
- Multi-rater Reconstruction
- Multi-rater Perception