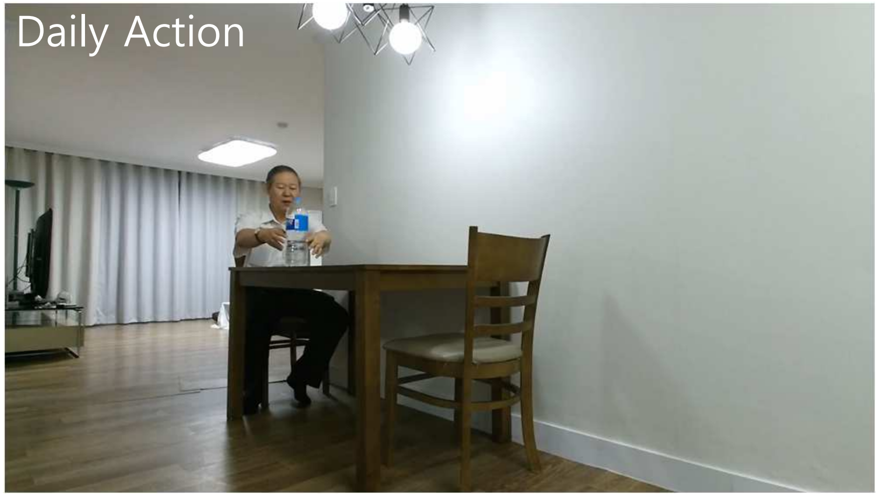
UST seminar

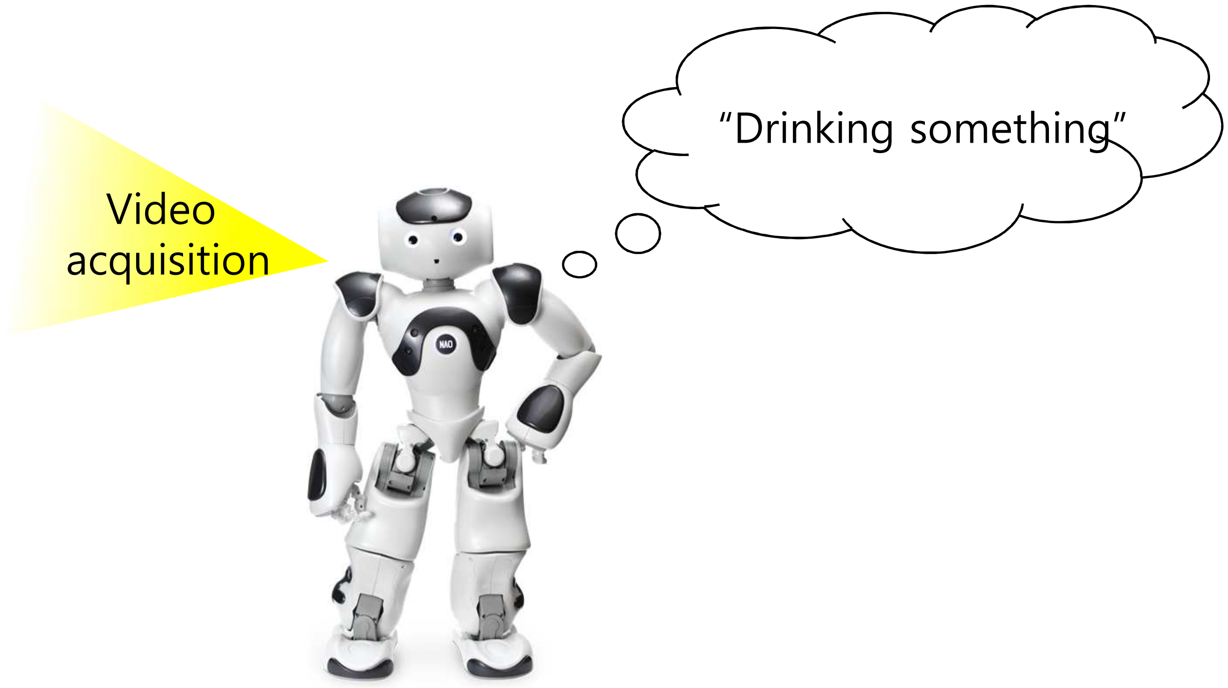
Ongoing result

Hyungmin Kim
UST-ETRI
Ph.D. student
khm159@etri.re.kr/ust.ac.kr
23th Nov, 2023

Human action recognition in Human-Robot Interaction



Video from [1]

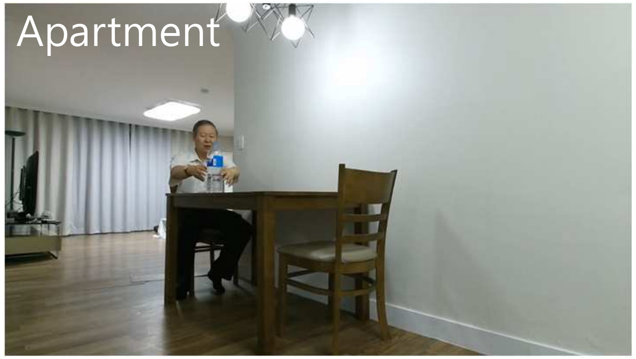


- **Problem definition :**

  Given input video v_i
  Model predicts corresponding action label y_i

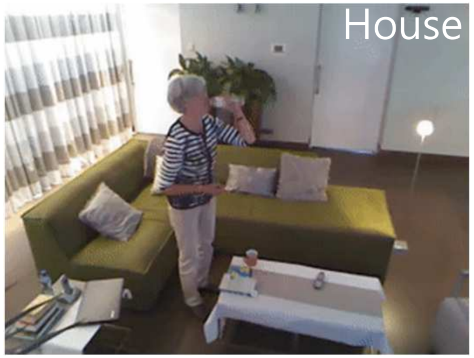
[1] Jang, Jinhyeok, et al. "ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly." *IROS*, 2020.

Service robot encounters numerous domains



Video from [1]

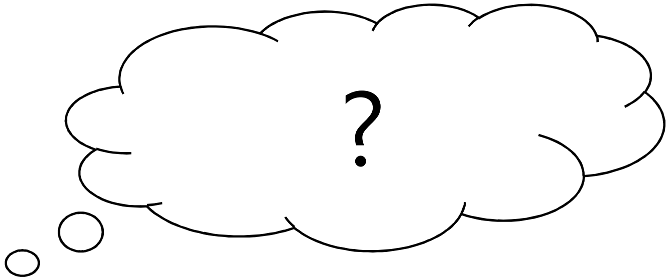
Video from [2]



Video from [3]

- Care-robot encounters numerous domains

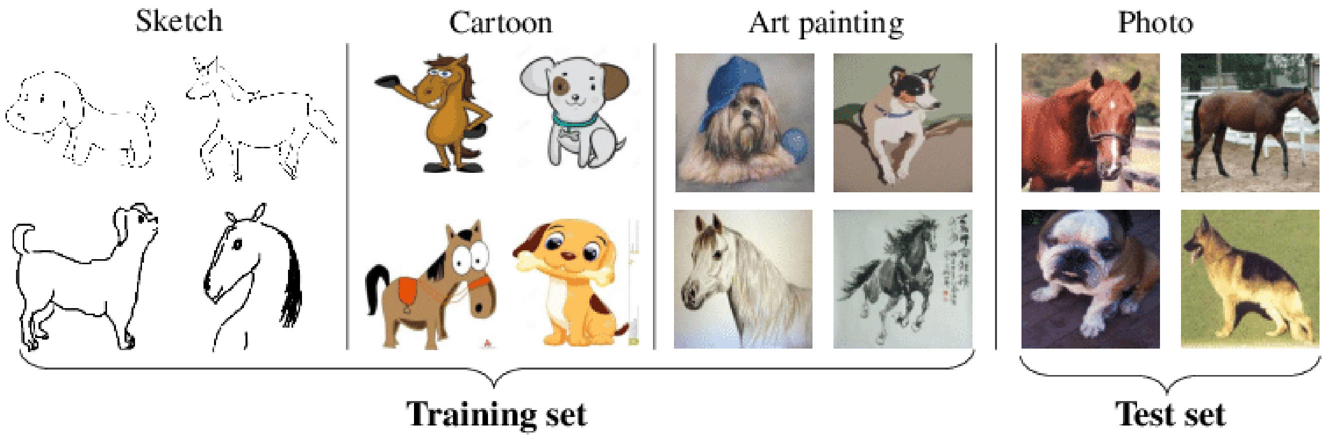
  Apartment, House, Building, Office ....



[1] Jang, Jinhyeok, et al. "ETRI-activity3D: A large-scale RGB-D dataset for robots to recognize daily activities of the elderly." *IROS*, 2020.
[2] Shahroudy, Amir, et al. "Ntu rgb+ d: A large scale dataset for 3d human activity analysis." *CVPR*. 2016.
[3] Das, Srijan, et al. "Toyota smarthome: Real-world activities of daily living." *ICCV*. 2019.

What is Domain Generalization?

Domain Generalization?



Train a generalized model with datasets from different domains that share the same semantic to improve performance on Un-seen domain.

[4] Wang, Jindong, et al. "Generalizing to unseen domains: A survey on domain generalization." *IEEE Transactions on Knowledge and Data Engineering* (2022).

UCF-HMDB benchmark

UCF-HMDB full dataset [1]

Fig. 3. Snapshots of all 12 categories of the cross-dataset UCF-HMDB benchmark. Spatial and temporal domain shifts co-exist in these scenarios.

- We select the previous UCF-HMDB_full benchmark
- The total of 2,009 videos colledted from the UCF-101 dataset and 1,800 videos sampled from HMDB51 datset
- There are 12 intersection classes.

[1] Chen, Min-Hung, et al. "Temporal attentive alignment for large-scale video domain adaptation." *ICCV,* 2019.

Motivation

Published as a conference paper at ICLR 2022

UNCERTAINTY MODELING FOR OUT-OF-DISTRIBUTION GENERALIZATION

Xiaotong Li[1], Yongxing Dai[1], Yixiao Ge[2], Jun Liu[3], Ying Shan[2], Ling-Yu Duan[1,4*]
[1]Peking University, Beijing, China [2]ARC Lab, Tencent PCG
[3]Singapore University of Technology and Design, Singapore
[4]Peng Cheng Laboratory, Shenzhen, China
lixiaotong@stu.pku.edu.cn, {yongxingdai, lingyu}@pku.edu.cn,
{yixiaoge, yingsshan}@tencent.com, jun_liu@sutd.edu.sg

Motivation

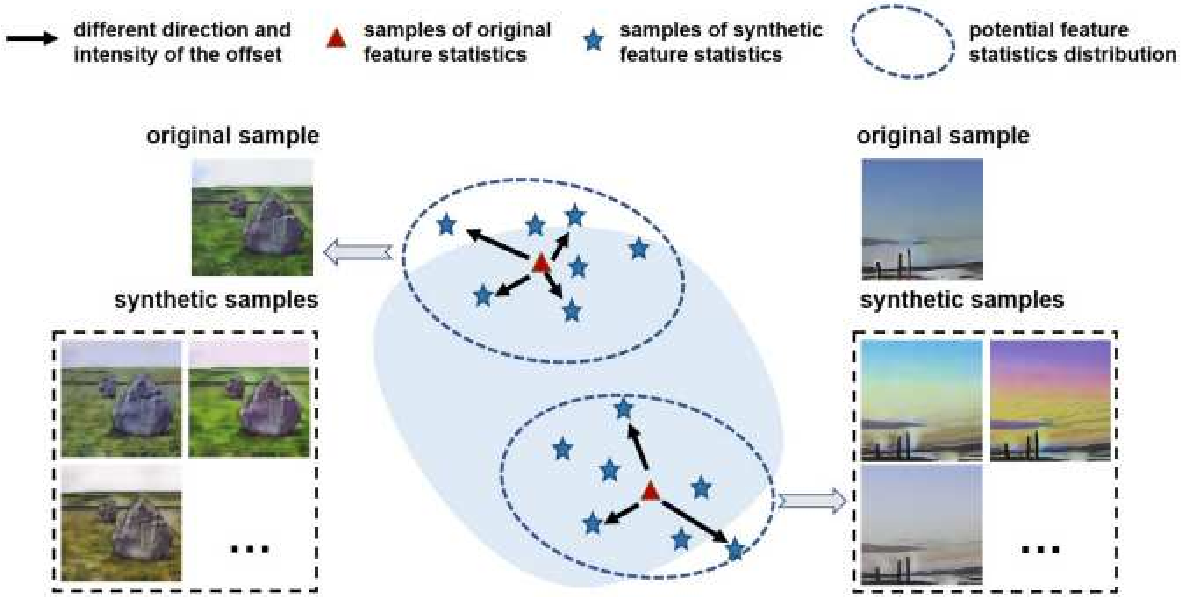


Figure 1: The visualization of reconstructed samples with synthesized feature statistics, using a pre-trained style transfer auto-encoder (Huang & Belongie, 2017). The illustration of the feature statistics shifts, which may vary in both intensity and direction (*i.e.*, different offsets in the vector space of feature statistics). We also show images of "new" domains generated by manipulating feature statistic shifts with different direction and intensity. Note these images are for visualization only, rather than feeding into the network for training.

Motivation

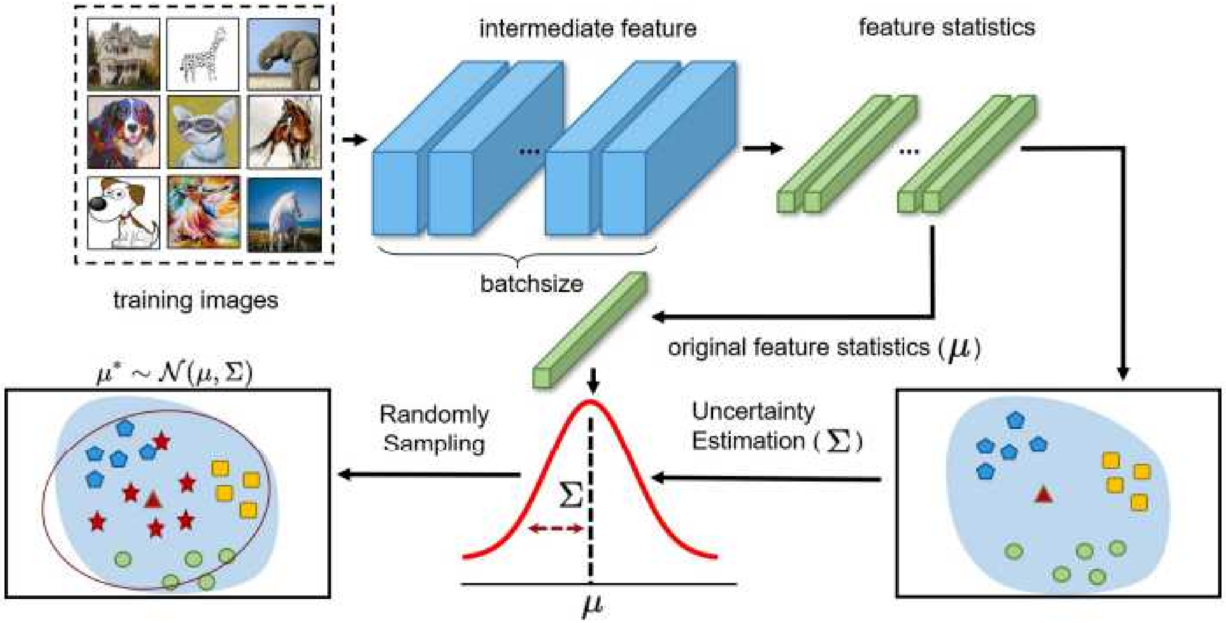
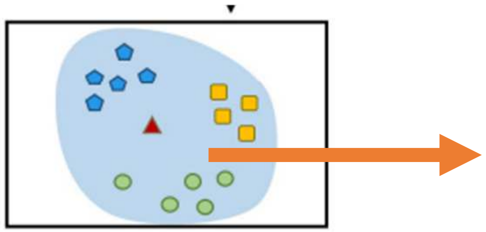


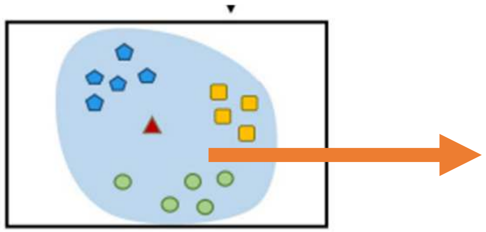
Figure 2: Illustration of the proposed method. Feature statistic is assumed to follow a multi-variate Gaussian distribution during training. When passed through this module, the new feature statistics randomly drawn from the corresponding distribution will replace the original ones to model the diverse domain shifts.

Motivation



The channel-wise μ and σ of
the intermediate features. samples
within a batch,

Motivation



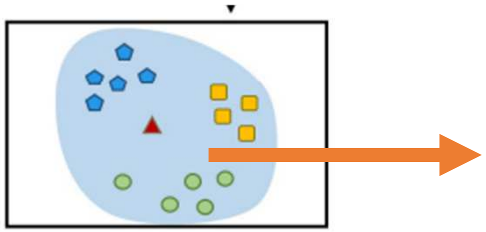
The channel-wise μ and σ of the intermediate features. samples within a batch,

분산 측정 (Uncertainty Estimation)

$$\Sigma_{\mu}^{2}(x) = \frac{1}{B}\sum_{b=1}^{B}(\mu(x) - \mathbb{E}_b[\mu(x)])^2,$$

$$\Sigma_{\sigma}^{2}(x) = \frac{1}{B}\sum_{b=1}^{B}(\sigma(x) - \mathbb{E}_b[\sigma(x)])^2.$$

Motivation

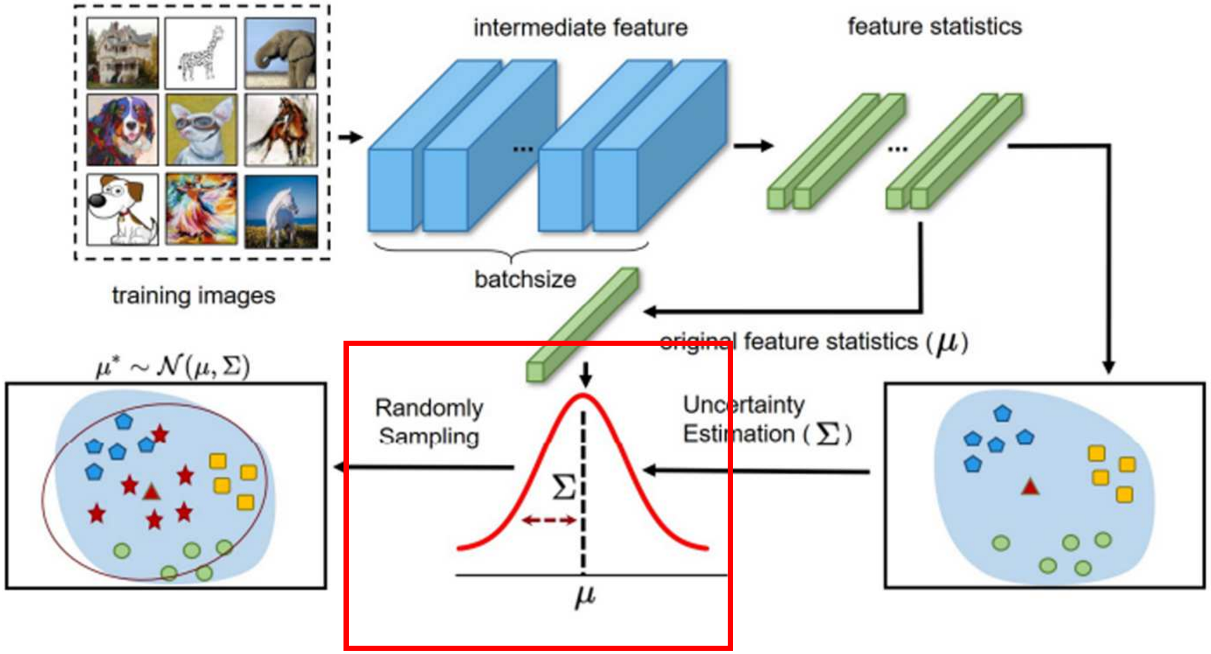


The channel-wise μ and σ of the intermediate features. samples within a batch

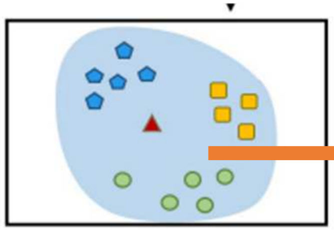
Affine transform 파라미터 샘플링

$$\beta(x) = \mu(x) + \epsilon_\mu \Sigma_\mu(x), \qquad \epsilon_\mu \sim \mathcal{N}(\mathbf{0}, \mathbf{1}),$$

$$\gamma(x) = \sigma(x) + \epsilon_\sigma \Sigma_\sigma(x), \qquad \epsilon_\sigma \sim \mathcal{N}(\mathbf{0}, \mathbf{1}).$$



Motivation



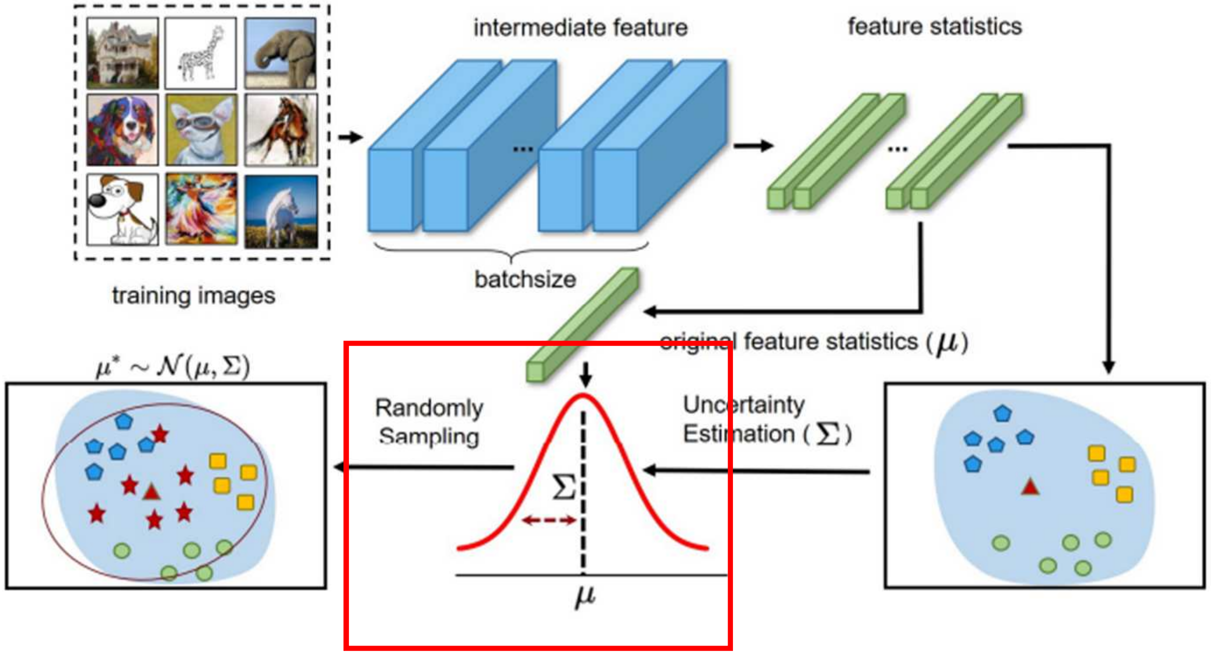
The channel-wise μ and σ of the intermediate features. samples within a batch

Affine transform 파라미터 샘플링

$$\beta(x) = \boxed{\mu(x)} + \boxed{\epsilon_\mu \Sigma_\mu(x)}, \qquad \epsilon_\mu \sim \mathcal{N}(\mathbf{0}, \mathbf{1}),$$

$$\gamma(x) = \boxed{\sigma(x)} + \boxed{\epsilon_\sigma \Sigma_\sigma(x)}, \qquad \epsilon_\sigma \sim \mathcal{N}(\mathbf{0}, \mathbf{1}).$$

original style rep. sampled uncertain(diversified) style rep. = perturbance

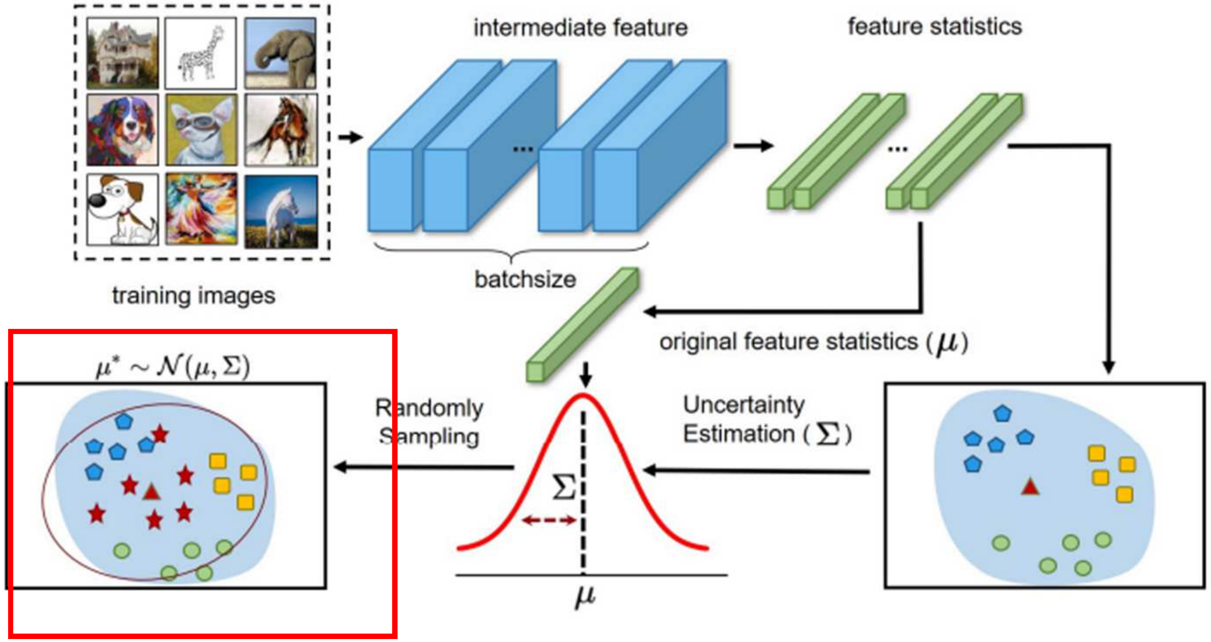


Motivation

$$\mathrm{DSU}(x) = \underbrace{(\sigma(x) + \epsilon_\sigma \Sigma_\sigma(x))}_{\gamma(x)} \left(\frac{x - \mu(x)}{\sigma(x)} \right) + \underbrace{(\mu(x) + \epsilon_\mu \Sigma_\mu(x))}_{\beta(x)}.$$

$$\gamma(x)(\frac{x - \mu(x)}{\sigma(x)}) + \beta(x)$$

Uncertainty를 고려한 style shifting 수행



Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

e)

Confidential

e)

Confidential

Confidential

e)

Confidential

e)

Confidential

Confidential

Confidential

e)

Confidential

e)

Confidential

e)

Confidential

*

** Lin, Kun-Yu, et al. "Diversifying Spatial-Temporal Perception for Video Domain Generalization." *NeurIPS, 2023*

e)

Confidential

*
** Lin, Kun-Yu, et al. "Diversifying Spatial-Temporal Perception for Video Domain Generalization." *NeurIPS, 2023*

e)

Confidential

*

** Lin, Kun-Yu, et al. "Diversifying Spatial-Temporal Perception for Video Domain Generalization." *NeurIPS, 2023*

Confidential

Confidential

*
** Lin, Kun-Yu, et al. "Diversifying Spatial-Temporal Perception for Video Domain Generalization." *NeurIPS, 2023*

Confidential

*

** Lin, Kun-Yu, et al. "Diversifying Spatial-Temporal Perception for Video Domain Generalization." *NeurIPS, 2023*

Confidential

Confidential

Confidential

Confidential

UST seminar

Towards domain-agnostic Video action recognition

Hyungmin Kim
UST-ETRI
Ph.D. student
khm159@etri.re.kr/ust.ac.kr
26th Oct, 2023