PECoP: Parameter Efficient Continual Pretraining for Action Quality Assessment

Amirhossein Dadashzadeh, Shuchao Duan, Alan Whone, Majid Mirmehdi
University of Bristol, UK

Overview

Problem

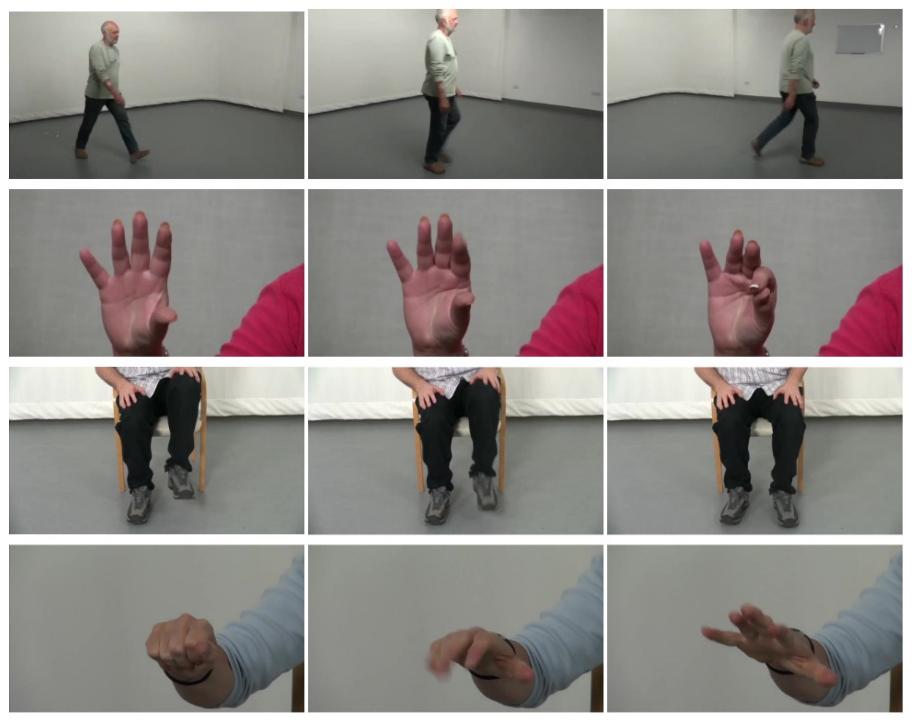
- Current AQA models face generalization issues due to limited labeled data and reliance on pretraining with general datasets, leading to significant domain shifts.

Key Contributions

I. We introduce PECoP, a novel workflow for parameter-efficient continual pretraining, enhancing the transfer of knowledge from large-scale datasets to AQA tasks and efficiently reducing the domain gap.

II. We integrate 3D-Adapter layer for the first time in 3D CNNs for video analysis.

III. We introduce a new annotated AQA dataset, PD4T, for the vision community to evaluate various actions performed by actual Parkinson's disease patients.

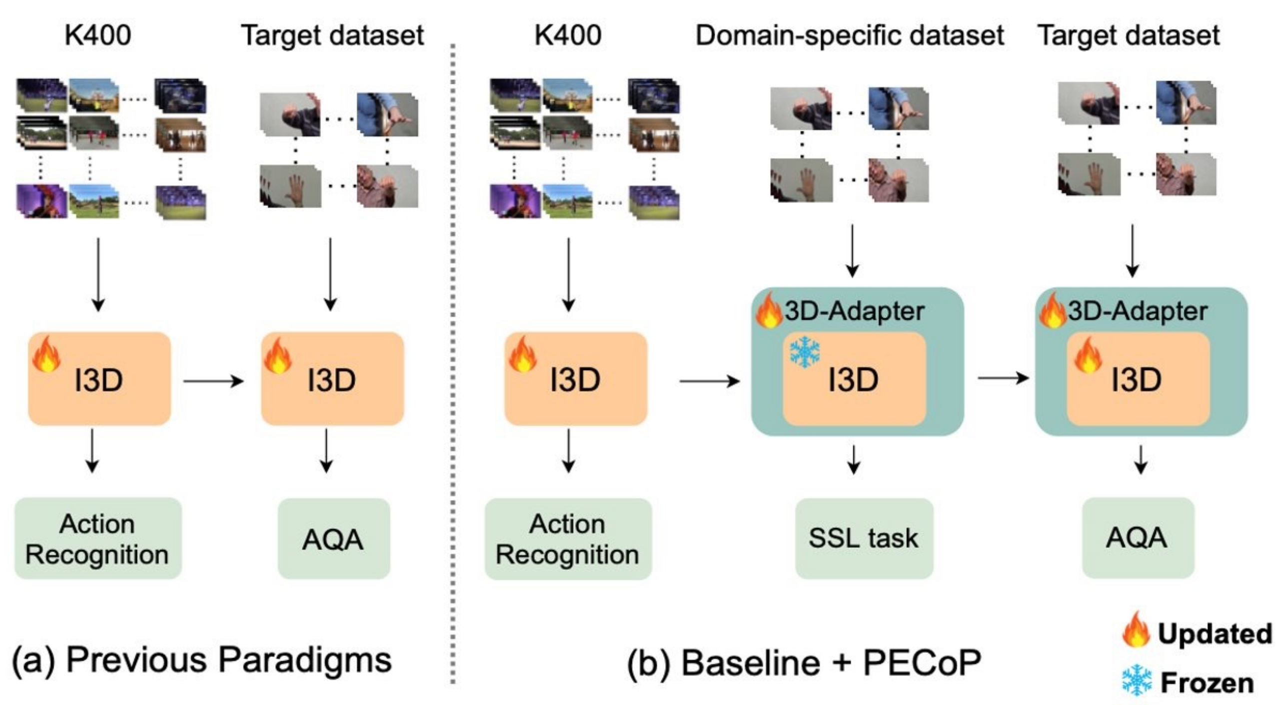


Sample frames from the PD4T dataset, from top to bottom, showing: gait, finger tapping, leg agility, and hand movement.

PECoP

PECoP vs Baseline

- Enhancing AQA with domain-specific pretraining.



(a) Previous Paradigms (b) Baseline + PECoP

Paper: arxiv.org/abs/2311.07603

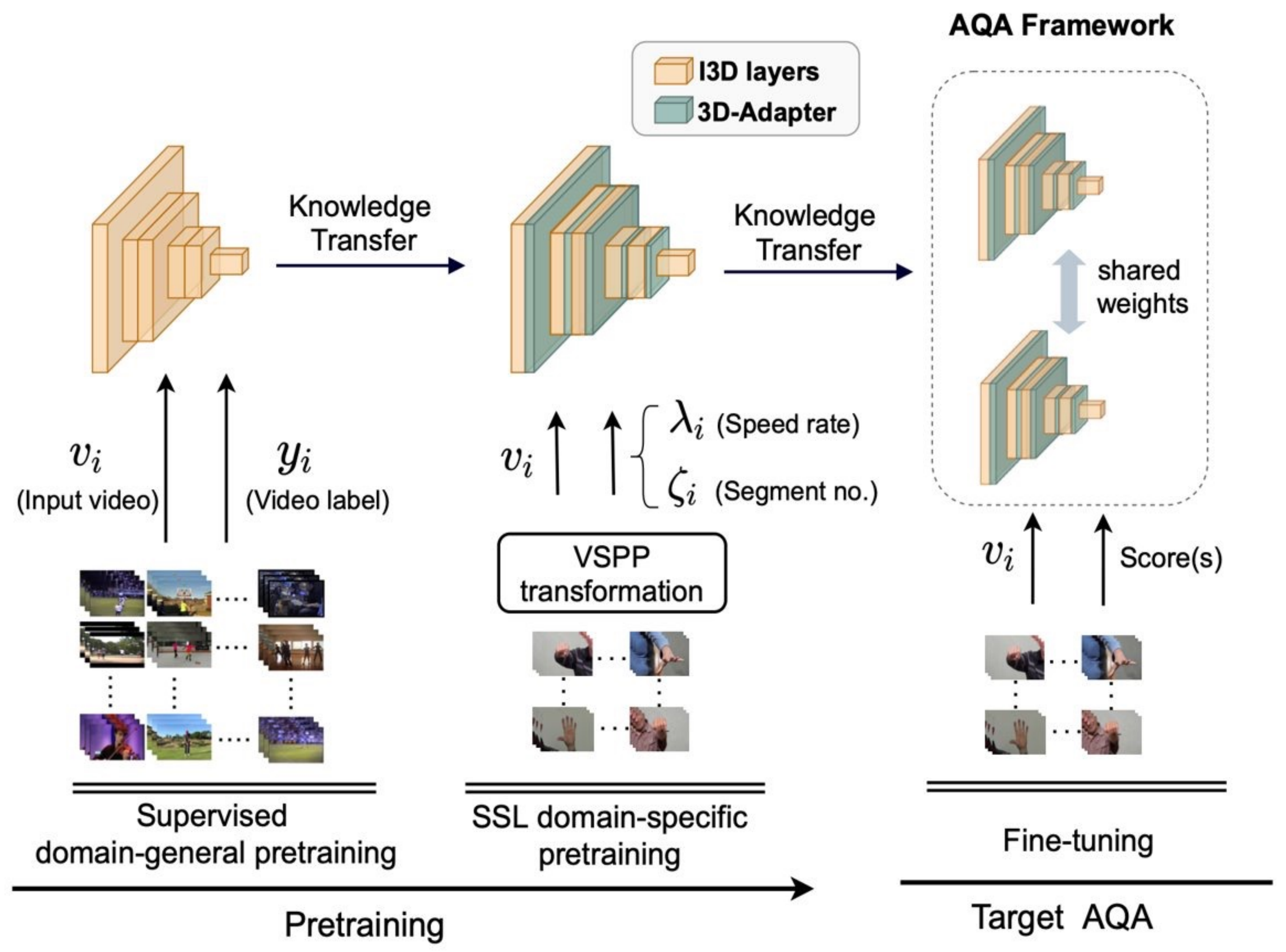
Code & Dataset: github.com/Plrbear/PECoP

Pretraining and Fine-tuning pipeline

I. Supervised pretraining on domain-general data e.g. Kinetics-400.

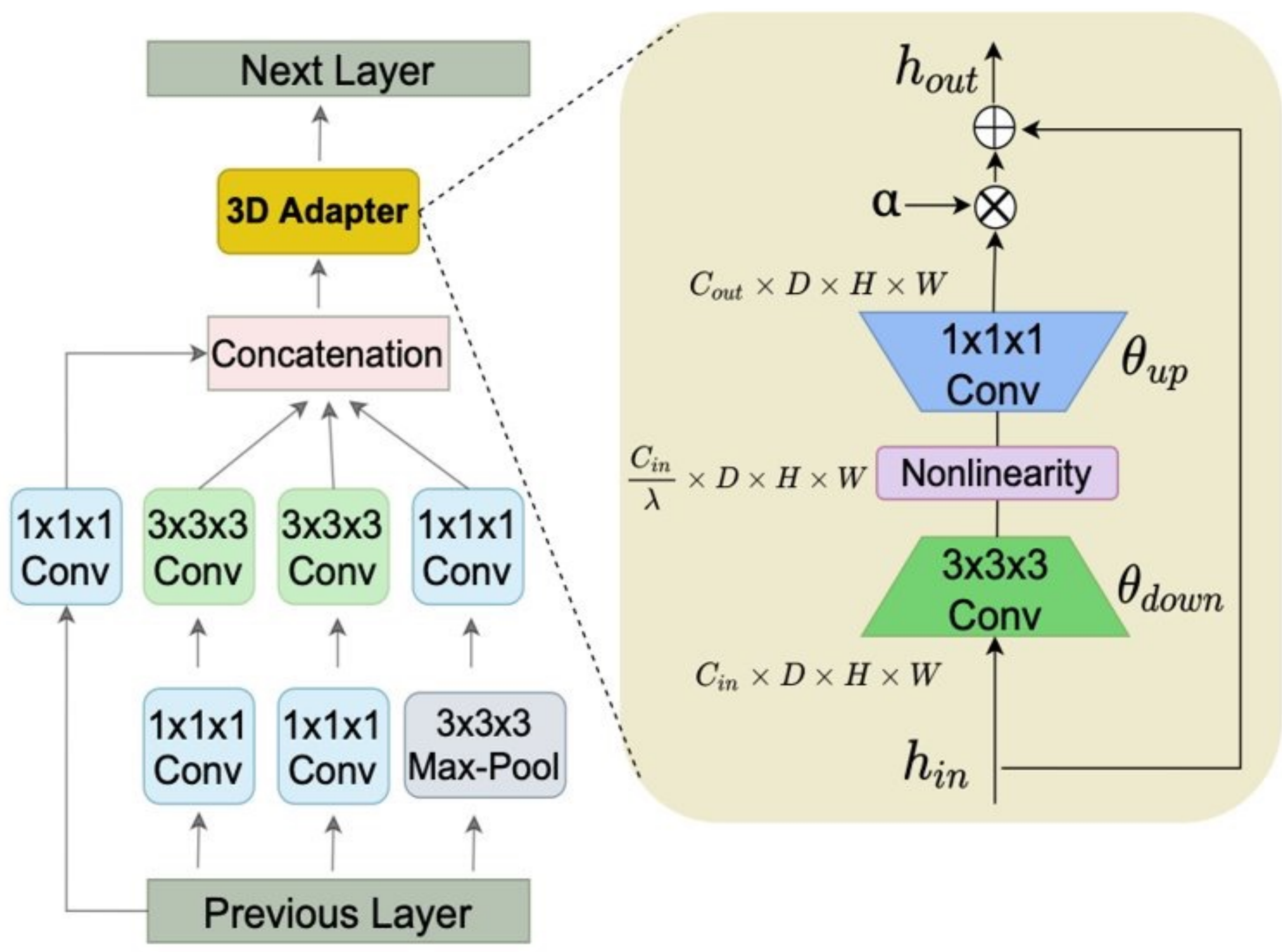
II. Self-supervised continual pretraining on domain-specific data e.g. target AQA dataset.

III. Fine-tuning the pretrained model on AQA target task using SOTA AQA methods (e.g. CoRe [I], USDL/MUSDL [2], and TSA [3]).



3D-Adapter

- The inception module (in I3D model) equipped with 3D-Adapter.



Experiments

Results

- Summary of Spearman Rank Correlation ($\mathcal{S}$) improvements for baseline methods with HPT [4] and PECoP.

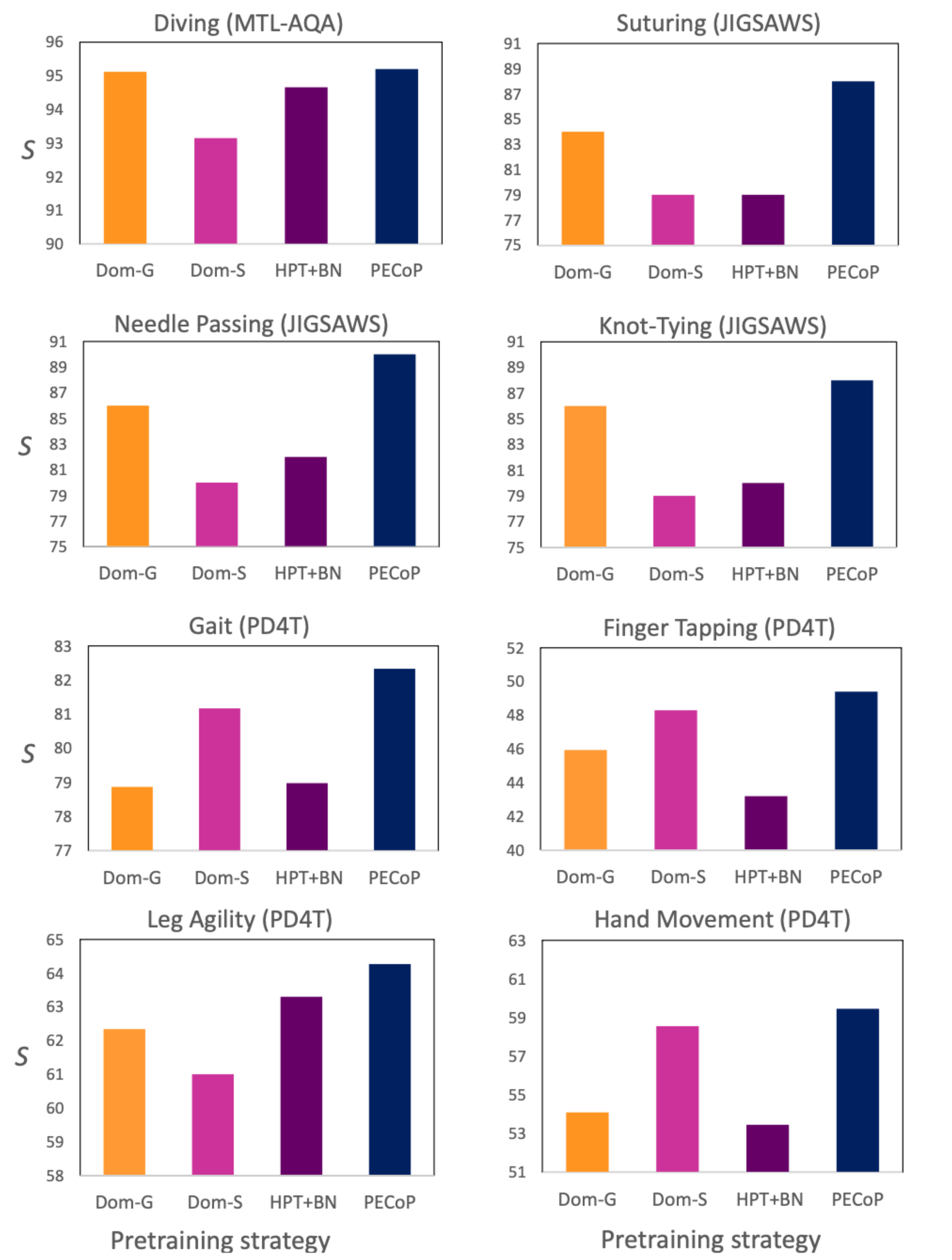
Dataset	Baseline	$\mathcal{S}$	+HPT	+PECoP
MTL-AQA [6]	MUSDL [2]	92.73	93.49 (↑0.76%)	93.72 (↑**0.99%**)
JIGSAWS [5] (Avg)	MUSDL [2]	70	72 (↑2%)	76 (↑**6%**)
JIGSAWS [5] (Avg)	CoRe [1]	85	80 (↓5%)	89 (↑**4%**)
PD4T (Avg)	CoRe [1]	60.31	63.05 (↑2.74%)	63.87 (↑**3.56%**)
FineDiving [3]	CoRe [1]	90.61	-	93.15 (↑**2.54%**)
FineDiving [3]	TSA [3]	92.03	-	93.13 (↑**1.1%**)

Efficiency (PECoP vs HPT)

Continual Pretraining	#trainble parameters	#epochs	Size
HPT [4]	~13M	16	~54MB
PECoP	~1M	8	~4MB

Ablation Study

- Comparison of PECoP with Domain-Specific SSL Pretraining (Dom-S), Domain-General Pretraining (Dom-G), and BatchNorm Tuning (HPT+BN) [4].



References

[1] Yu, Xumin, et al. "Group-aware contrastive regression for action quality assessment." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

[2] Tang, Yansong, et al. "Uncertainty-aware score distribution learning for action quality assessment." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

[3] Xu, Jinglin, et al. "Finediving: A fine-grained dataset for procedure-aware action quality assessment." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.

[4] Reed, Colorado J., et al. "Self-supervised pretraining improves self-supervised pretraining." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

[5] Gao, Yixin, et al. "Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling." *MICCAI workshop: M2cai*. Vol. 3. No. 3. 2014.

[6] Parmar, Paritosh, and Brendan Tran Morris. "What and how well you performed? a multitask learning approach to action quality assessment." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.