Australian National University

Adversarial Training based HRNet (HRGAN)

Group 15

Presenter:

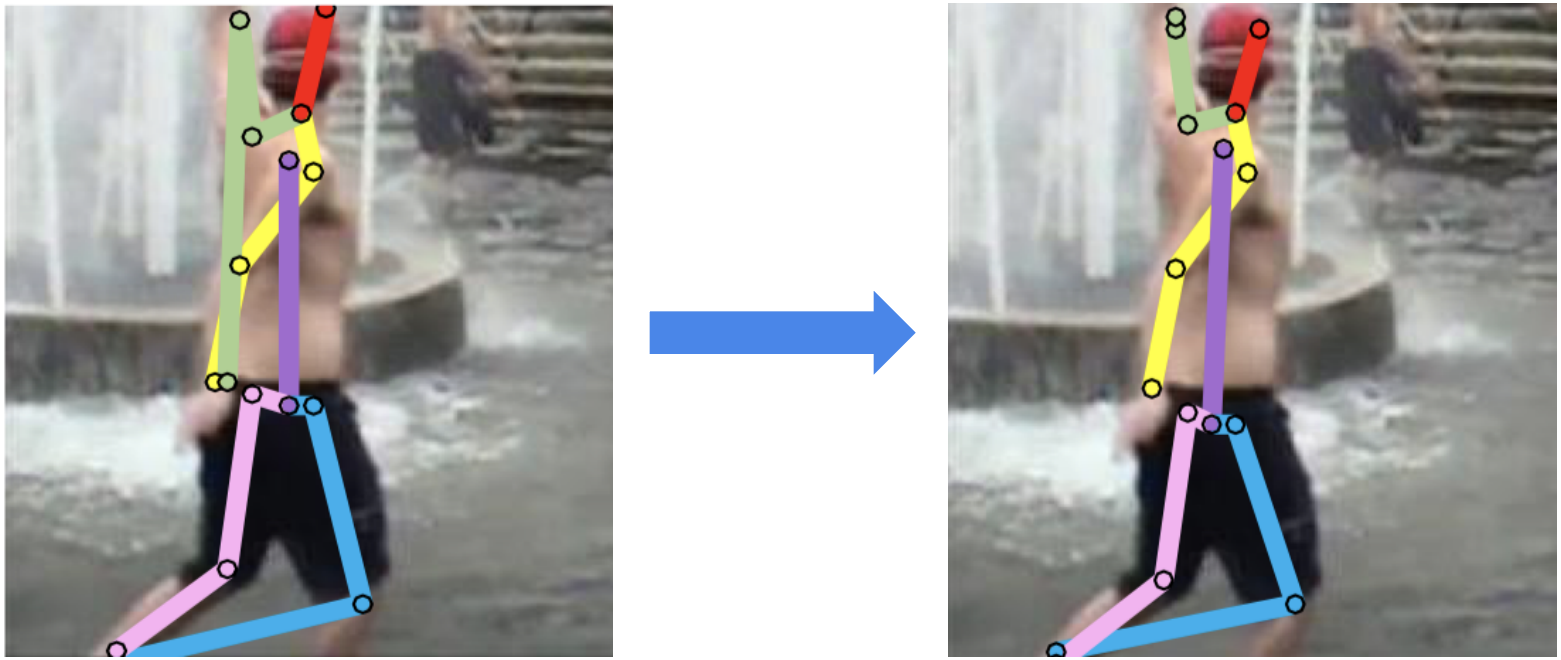
Ting Cao (u7078470)

Siyuan Yan (u7050317)

Gefei Zhang (u7138112)

Luning Li (u7077148)

Challenge—Implausible pose

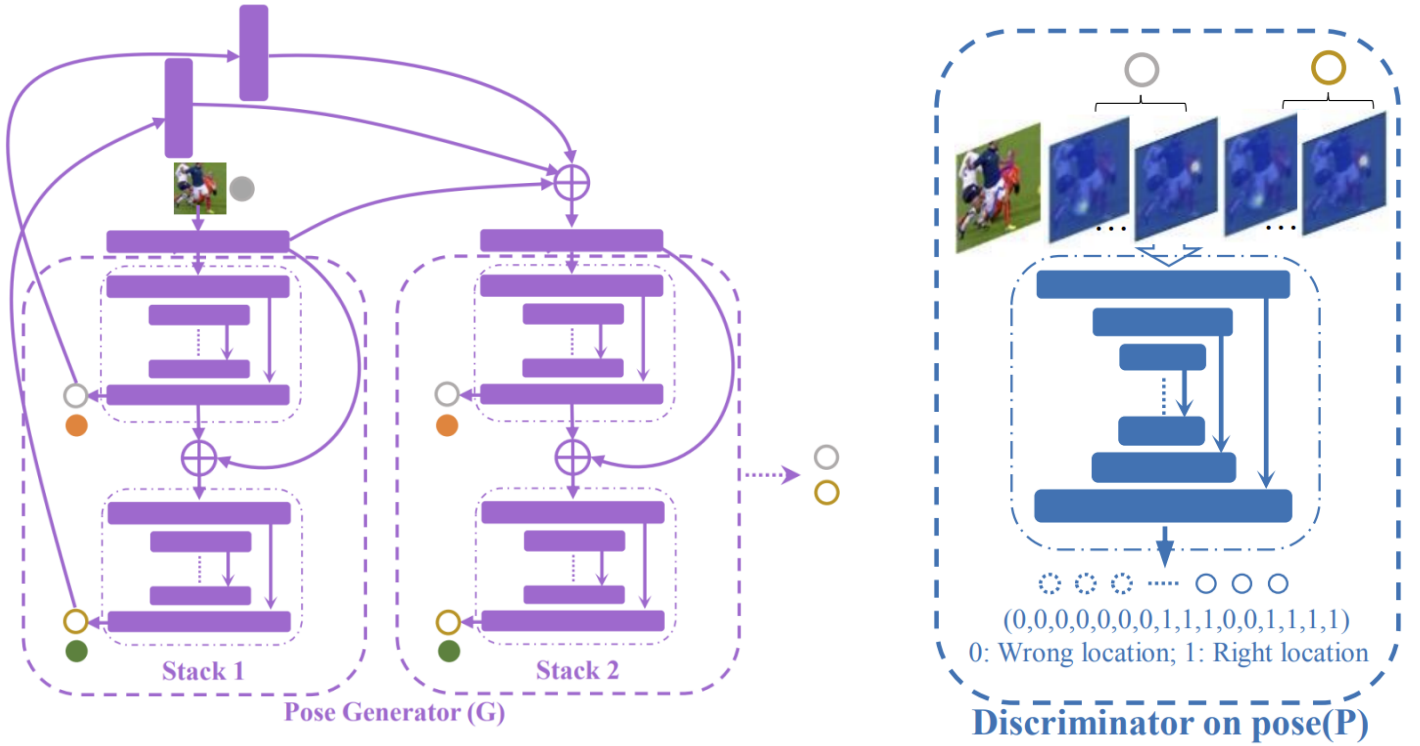


Challenge—Occlusion

Objective

- To prove the effectiveness of adversarial training on HRNet
- To solve the occlusion problem in crowded scene
- To correctly assign implausible pose

Previous work—GAN



:Predicted pose heatmap

: Groundtruth pose heatmap

:Predicted occlusion heatmap

:Groundtruth occlusion heatmap

Previous work—HRnet

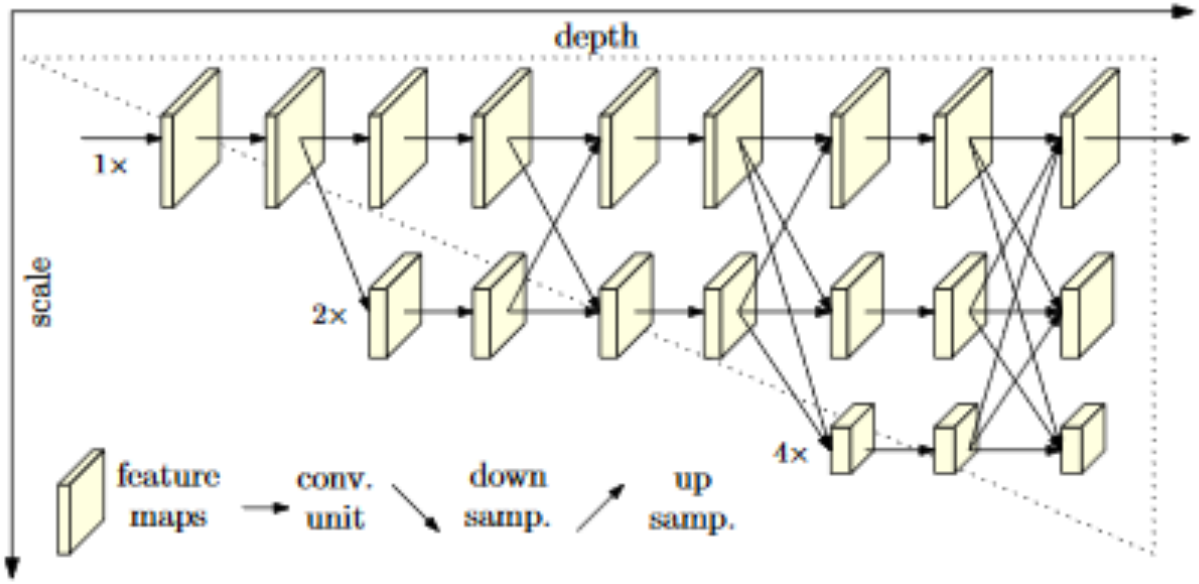


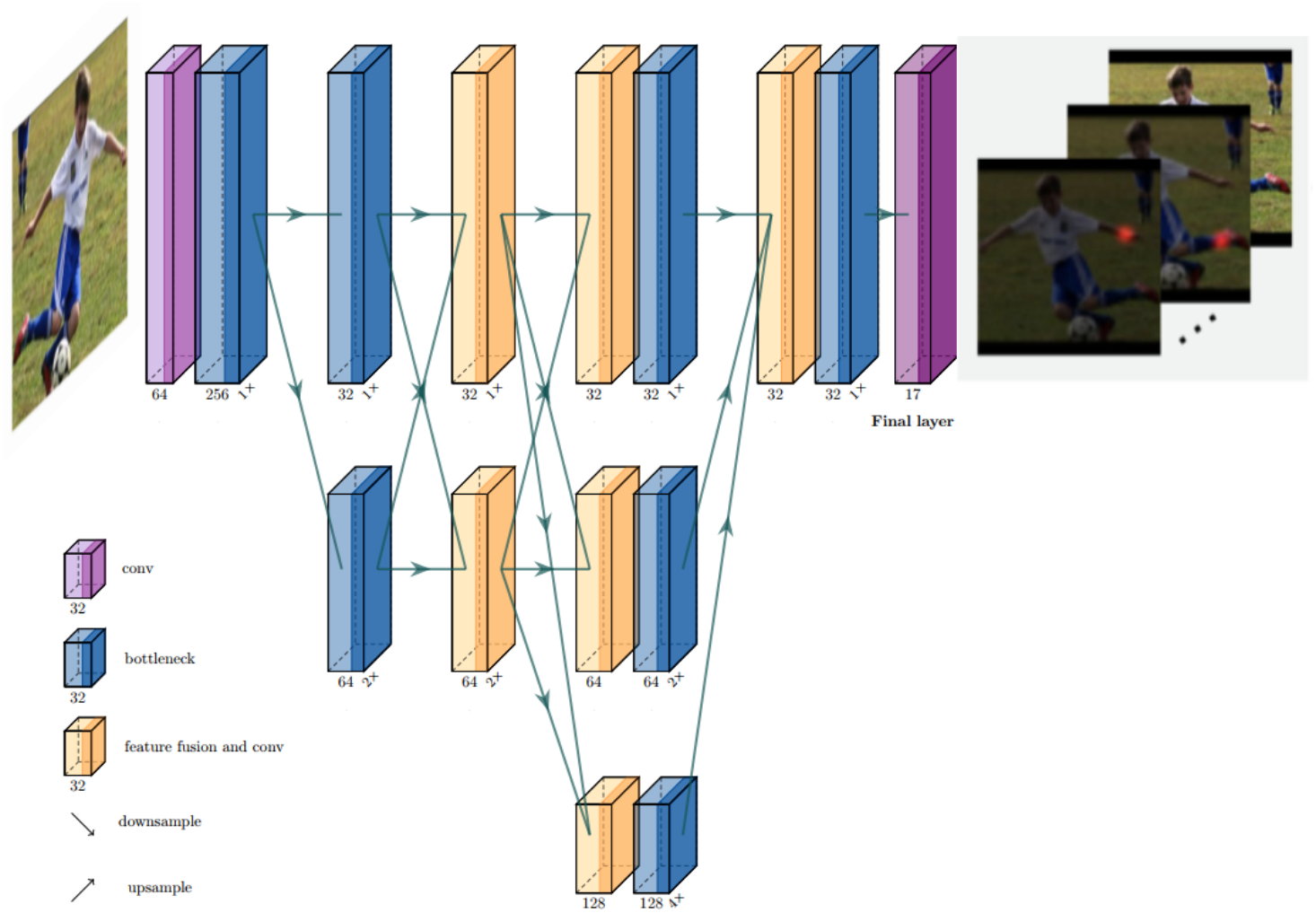
Figure 1. Illustrating the architecture of the proposed HRNet. It consists of parallel high-to-low resolution subnetworks with repeated information exchange across multi-resolution subnetworks (multi-scale fusion). The horizontal and vertical directions correspond to the depth of the network and the scale of the feature maps, respectively.

Previous work—BEGAN

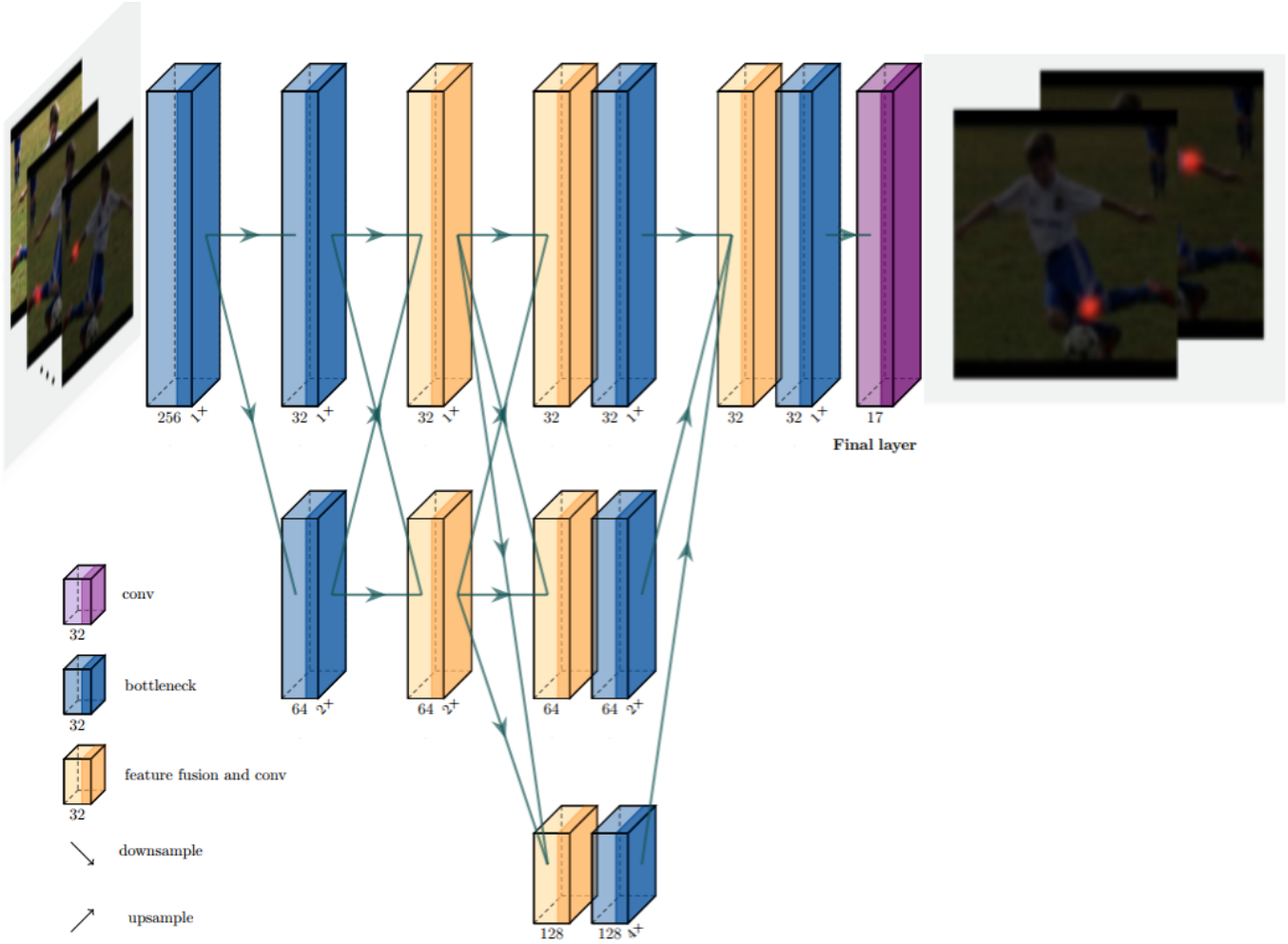
Boundary equilibrium

$$\begin{cases} \mathcal{L}_D = \mathcal{L}(x) - k_t.\mathcal{L}(G(z_D)) & \text{for } \theta_D \\ \mathcal{L}_G = \mathcal{L}(G(z_G)) & \text{for } \theta_G \\ k_{t+1} = k_t + \lambda_k(\gamma\mathcal{L}(x) - \mathcal{L}(G(z_G))) & \text{for each training step } t \end{cases}$$

Architecture—Generator



Architecture—Discriminator



Training Strategy for Generator

content loss and adversarial loss

$$\mathcal{L}_G = \mathcal{L}_{\mathcal{MSE}} + \lambda_G \mathcal{L}_{adv}$$

$$\mathcal{L}_{MSE} = \sum_{j=1}^{M} (C_j - \hat{C}_j)^2$$

$$\mathcal{L}_{adv} = \sum_{j=1}^{M} (C_j - D(\hat{C}_j, X))^2$$

Training Strategy for Discriminator

$$\mathcal{L}_{\text{real}} = \sum_{j=1}^{M} (C_j - D(C_j, X))^2,$$

$$\mathcal{L}_{\text{fake}} = \sum_{j=1}^{M} (\hat{C}_j - D(\hat{C}_j, X))^2,$$

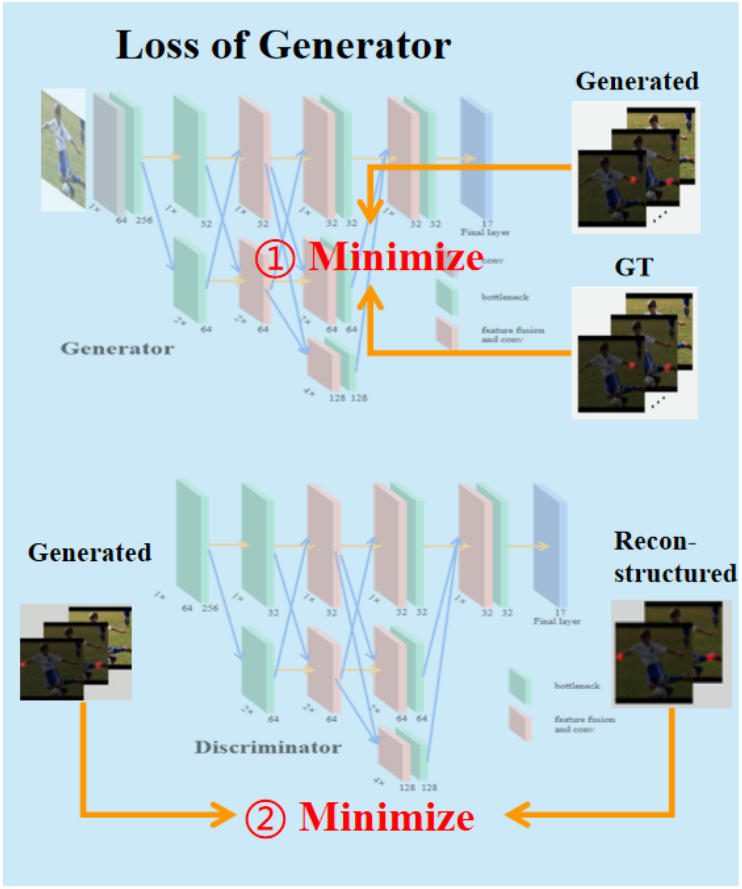
$$\mathcal{L}_D = \mathcal{L}_{\text{real}} - k_t \, \mathcal{L}_{\text{fake}} \, .$$

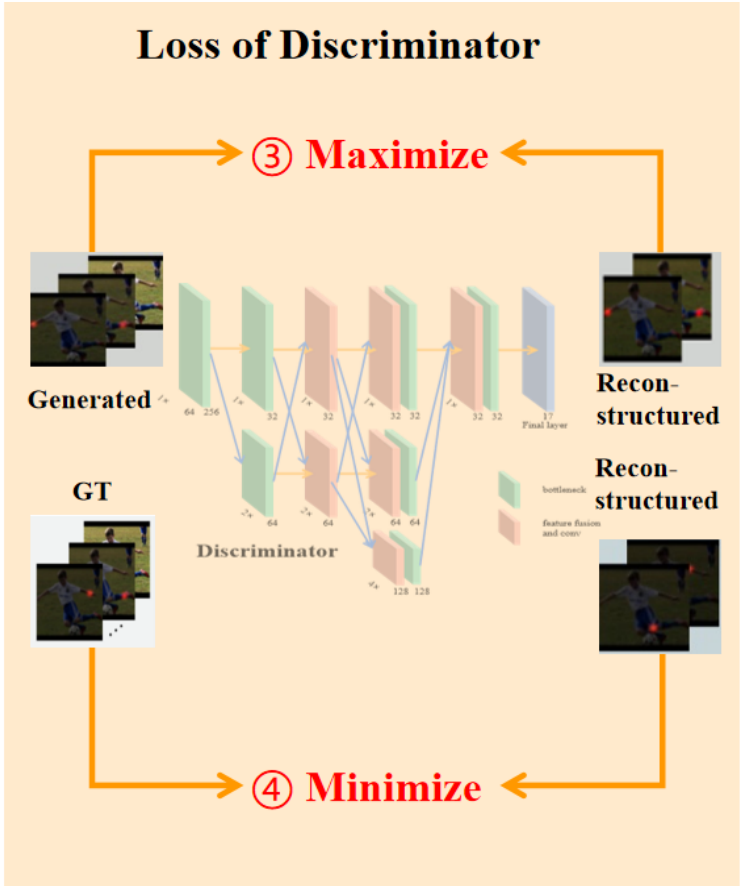
Boundary Equilibrium

$$k_{t+1} = k_t + \lambda_k \left(\gamma \mathcal{L}_{\text{real}} - \mathcal{L}_{\text{fake}} \right)$$

- To control the balance between generator and discriminator.
- To solve the mode collapse

Adversarial Training Process





Experiment Results

MpII trained on 1 V100 GPU, ~90 epochs
COCO trained on 4 V100 GPUs, ~130 epochs

MPII

Arch	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean
Self Adversarial Training (hourglass)	98.2	96.8	92.2	88.0	91.3	89.1	84.9	91.8
Adversarial PoseNet (hourglass)	98.6	96.4	92.4	88.6	91.5	88.6	85.7	92.1
hrnet	**98.6**	96.9	**92.8**	**89.0**	91.5	89.0	**85.7**	92.3
hrgan (ours)	97.8	**97.5**	92.5	87.5	**93.9**	**89.7**	85.1	**92.5**

COCO

Arch	AP	Ap .5	AP .75	AP (M)	AP (L)	AR	AR .5	AR .75	AR (M)	AR (L)
hrnet	**0.744**	0.905	0.819	0.708	**0.810**	0.798	0.942	0.865	0.757	**0.858**
hrgan (ours)	0.742	**0.906**	0.819	**0.711**	0.807	**0.801**	**0.946**	**0.869**	**0.761**	0.857

Ablation study

Arch	Head	Shoulder	Elbow	Wrist	Hip	Knee	Ankle	Mean
HRGAN (w. boundary equilibrium)	98.0	**97.1**	**94.2**	89.6	**92.2**	**88.4**	86.1	**92.8**
hrnet+conventional GAN	- (collapsed)	-	-	-	-	-	-	-
hrnet+channel wise conventional GAN	- (collapsed)	-	-	-	-	-	-	-
w. final stage fusion	**98.3**	96.9	93.7	**90.1**	91.2	88.4	**86.3**	92.4
w. target wight	97.1	95.2	90.9	86.7	89.6	88.1	80.1	90.8

Occlusion

HRNet

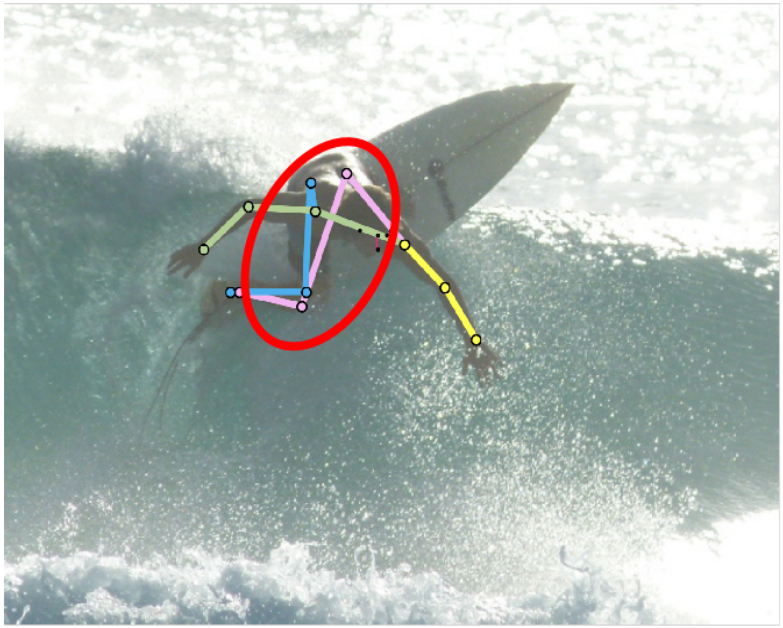
Our HRGAN

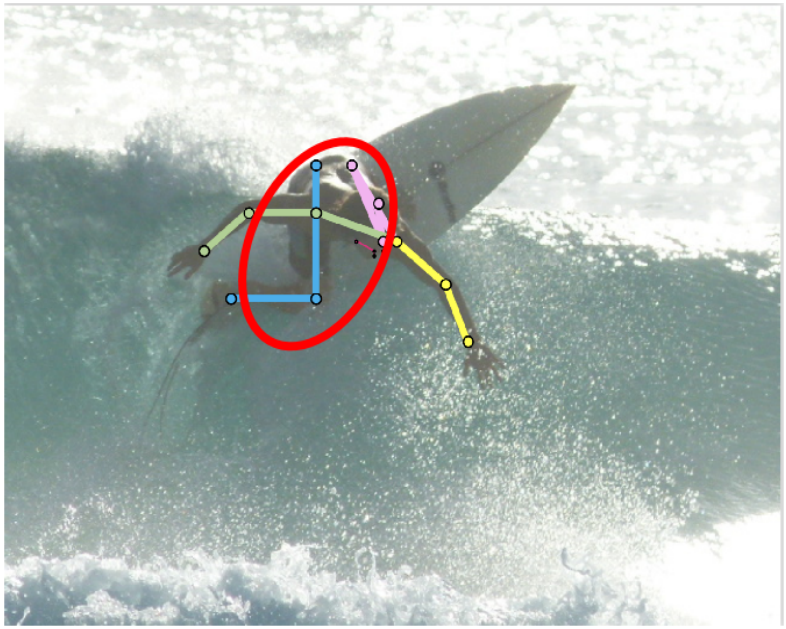
Occlusion

HRNet

Our HRGAN

Implausible Pose

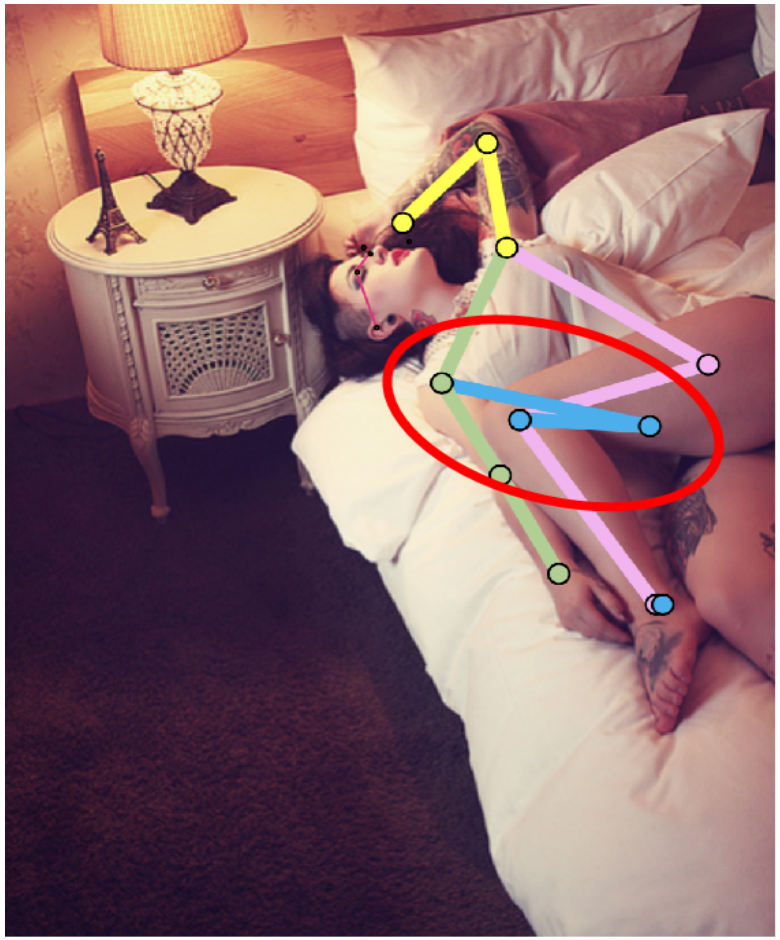




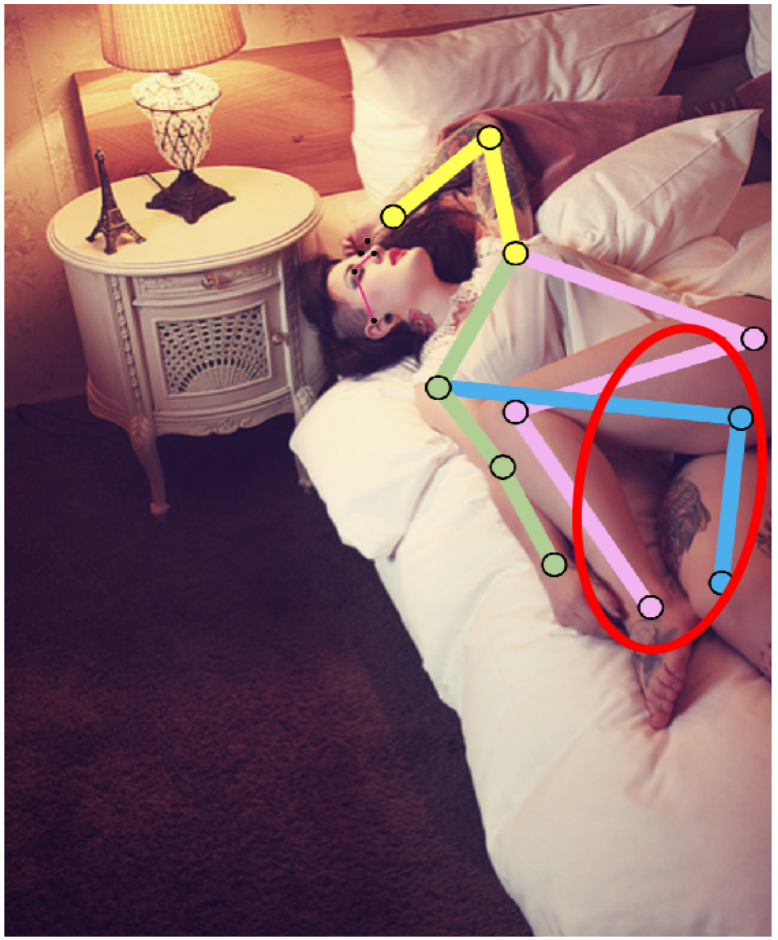
HRNet

Our HRGAN

Implausible Pose



HRNet



Our HRGAN

Limitation

- Performance is limited to its detection model
- High computational cost

Conclusion

We proposed an adversarial training based HRNet called HRGAN.

We solved the implausible pose and occlusion problem in HRNet.

We outperformed state-of-the-art top-down model HRNet in MPII.

Future Work

- Explore methods which can detect small person and low resolution
- Extend to bottom up method


Q&A