Learning Privacy-preserving Optics For Human Pose Estimation

High Dimensional Signal Processing Research Group

carlos.hinojosa@saber.uis.edu.co

Carlos Hinojosa[1], Juan Carlos Niebles[2], Henry Arguello[1]

[1]Universidad Industrial de Santander [2]Stanford University

STANFORD VISION & LEARNING

Motivation

Cameras are everywhere! How to develop privacy-preserving vision systems?

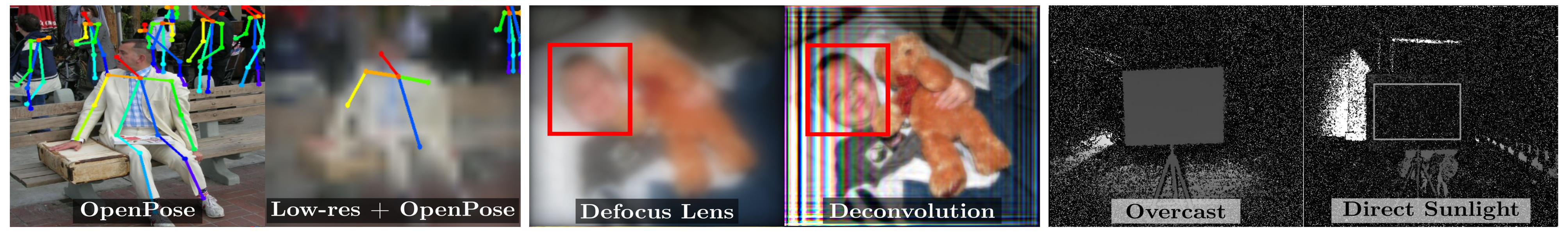
We want to prevent the camera from obtaining detailed visual data that may contain private information, desirably at the hardware level.

Prior work on Privacy-preserving vision

Low-resolution
- Lose information.
- Pose estimation fails.

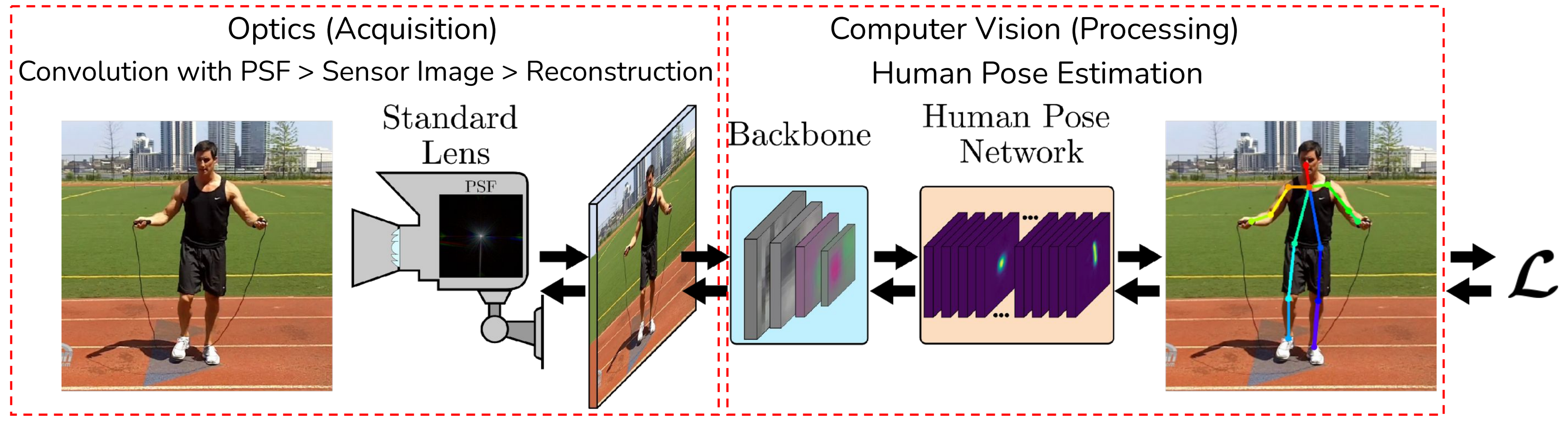
De-focusing
- Susceptible to reverse engineering attacks.

Depth cameras
- Bright sunlight degrades depth estimation quality.



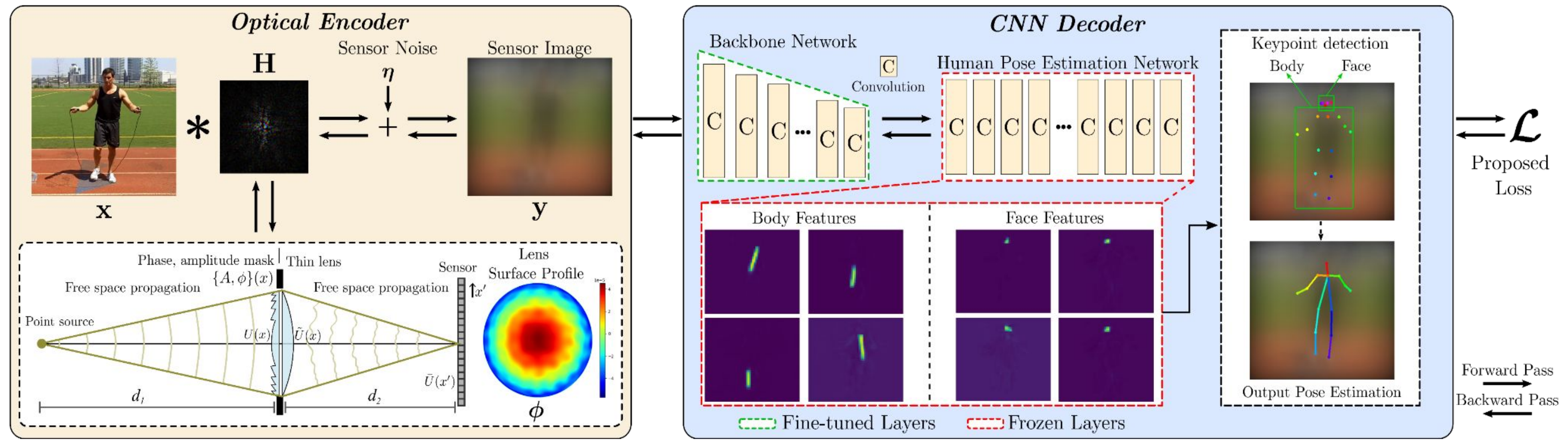
Our key idea: instead of fixed/manually define optics, we'll design optical distortion in a way that doesn't degrade the vision algorithm performance.

Traditional Deep-optics-based Computational Cameras



- The concept of *Deep Optics* refers to the joint design of optics and algorithms to boost the performance of the final task.
- All Deep Optics methods rely on the same approach: to remove the aberrations from the lens to obtain high-quality reconstructed images.

Model and Approach



- We rely on the converse approach of deep optics: We add aberrations to the lens to obtain privacy protection and jointly perform HPE.
- Our optimization process has two parts: an optical encoder, which provides hardware-level privacy protection by degrading the image quality, and a CNN decoder that learns features from the highly degraded images to perform HPE.

End-to-end Optimization

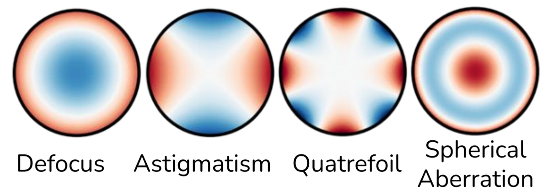
Formally, we formulate our optimization problem by combining the two goals: to acquire privacy-preserving images and to perform HPE with high accuracy.

$$\alpha^*, h^* = \arg\min_{\alpha,h} L_T(h) + L_P(\alpha).$$

Lens Parametrization (α)

- We parameterize the surface profile of the lens with Zernike polynomials, where each one describes a wavefront aberration.

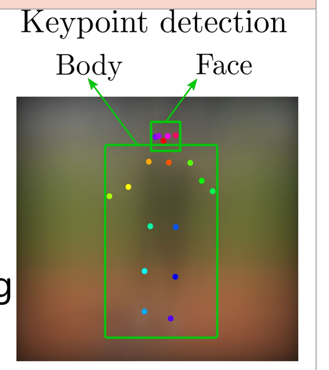
$$\phi = \sum_{j=1}^{q} \alpha_j \mathbf{Z}_j,$$



- We learn α_j
- ϕ Is the lens surface.

Human Pose Estimation Network (h)

- To perform HPE, we adopted the OpenPose (OPPS) network.
- We separate the face and body keypoints.
- We seek a network that accurately detects the body points while ignoring the face points.



Datasets and Metrics

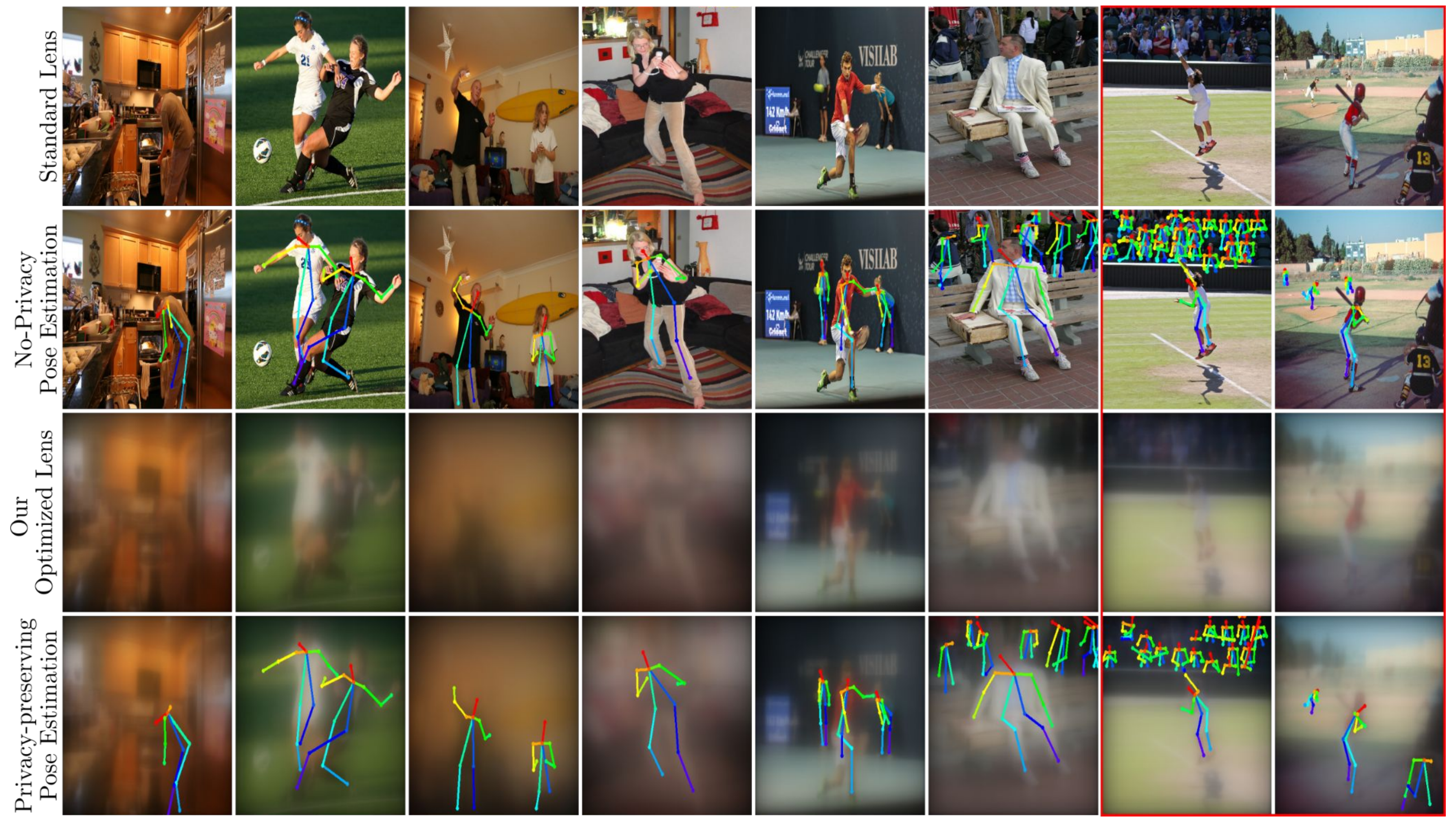
Dataset

We train our proposed end-to-end approach on the COCO 2017 keypoints dataset and evaluate our approach on the val2017 set.

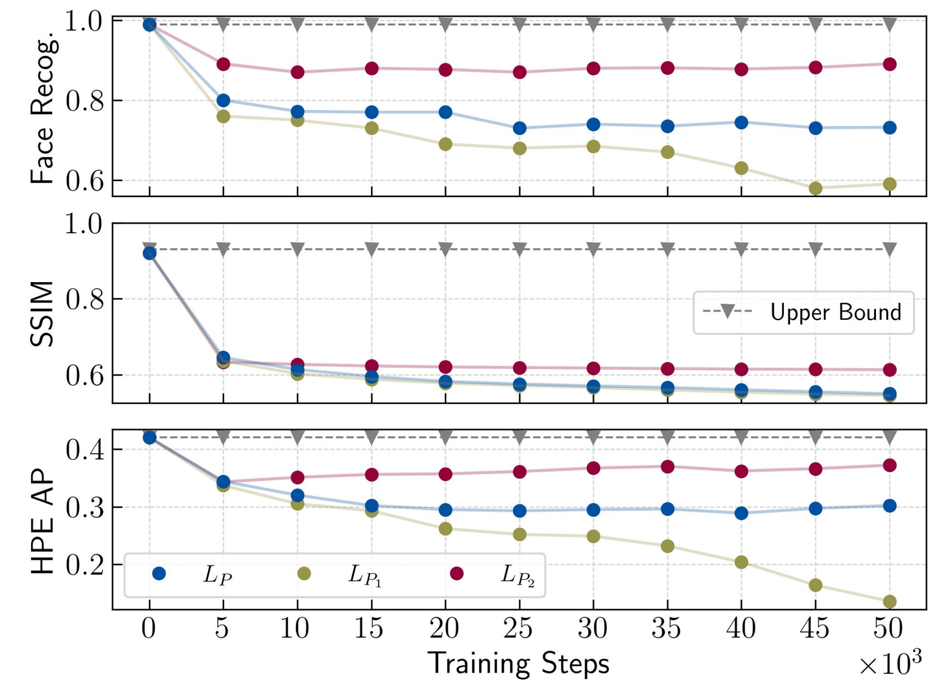
Metrics

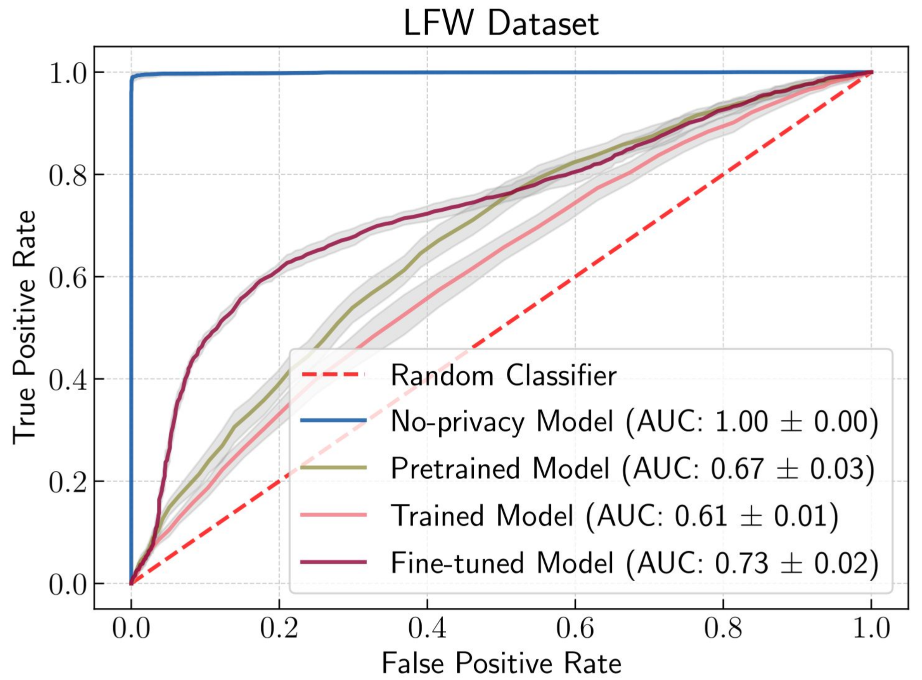
HPE	Face Recognition	Image Quality
We use the standard COCO evaluation metric: Object Keypoint Similarity (**OKS**). To make a fair comparison, we sightly modify the COCO evaluation script to not consider the face keypoints.	We implement the **ArcFace** network to measure privacy. We train ArcFace on three face recognition datasets. We measure its performance in terms of the area under the curve (**AUC**) of the **ROC**.	To measure image degradation, we use the peak-signal-to-noise ratio (**PSNR**) and the structural similarity index measure (**SSIM**). We expect to achieve lower PSNR and SSIM values.

Qualitative Results on Example COCO Images



Experiments: Ablation Studies





Quantitative Experiments: Comparison with Prior Works

Method	PSNR	SSIM	AP	AR
OPPS (Upper Bound)	-	-	0.421	0.506
Defocus Lens	16.614	0.598	0.197	0.256
Low-Resolution	18.54	0.476	0.067	0.106
PP-OPPS (Ours)	**14.851**	**0.567**	**0.302**	**0.363**

We compare our method against two traditional privacy-preserving approaches: Defocus and Low-resolution cameras. OPPS stands for the original OpenPose network. The PP prefix stands for our proposed approach.