

2025 하계 세미나

Practical 2D Human Pose Estimation



Sogang University

Vision & Display Systems Lab, Dept. of Electronic Engineering



Presented By

변재현

Outline

- Background
 - Human pose estimation
 - Evaluation metric
- DynPose: Largely Improving the Efficiency of Human Pose Estimation by a Simple Dynamic Framework
 - CVPR 2025
- ProbPose: A Probabilistic Approach to 2D Human Pose Estimation
 - CVPR 2025

Background

- Human pose estimation

- 개념

- 이미지 또는 영상에서 사람의 관절을 검출하여 2D 또는 3D 스켈레톤 구조로 표현

- 구현 방식

- Top-down approach

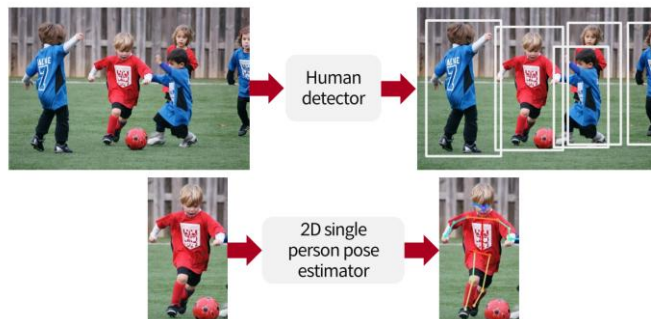
- ※ 사람의 BBOX를 detection하고, BBOX 내부에서 자세를 예측하는 방식

- ※ Bottom-up approach에 비해 정확도가 높지만, 그만큼 연산량이 많고 속도가 느림

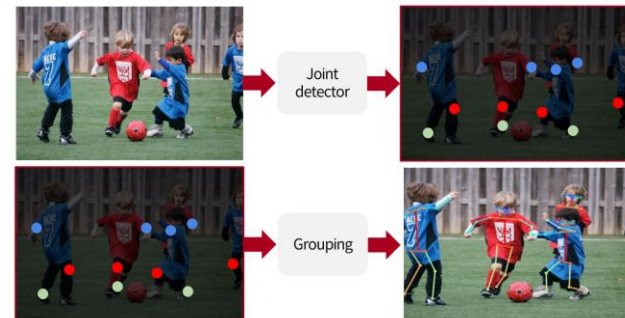
- Bottom-up approach

- ※ Keypoint를 먼저 예측하고, keypoint 간의 관계를 분석해서 자세를 예측하는 방식

- ※ Top-down approach에 비해 연산량이 적어서 빠르지만 정확도가 떨어짐



Top-down approach



Bottom-up approach

Background

- Evaluation metric
 - OKS (Object Keypoint Similarity)

$$OKS = \frac{\sum_i \exp\left(-\frac{d_i^2}{2s^2k_i^2}\right) \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$

d_i : 예측된 keypoint와 GT keypoint 사이의 유클리드 거리
 s : 해당 사람의 스케일
 k_i : 각 joint마다 설정된 상수
 $\delta(v_i > 0)$: 해당 keypoint의 가용성

- PCK

$$PCK = \frac{1}{N} \sum_{i=1}^N \mathbf{1}\left(\frac{\|\hat{\mathbf{p}}_i - \mathbf{p}_i\|_2}{L} \leq \alpha\right)$$

$\hat{\mathbf{p}}_i$: 예측 keypoint 좌표
 $\mathbf{p}_i$: GT keypoint 좌표
 L : 기준 길이 (예: bounding box 대각선 길이)
 α : 임계값 (예: 0.1 등)
 $\mathbf{1}(\cdot)$: 조건을 만족하면 1, 아니면 0

DynPose: Largely Improving the Efficiency of Human Pose Estimation by a Simple Dynamic Framework

[CVPR 2025]

DynPose¹⁾

- Introduction

- Problem statement

- Top-down HPE(Human Pose Estimation)은 정확하지만 연산 효율성 개선 연구 부족
 - 최근 연구는 모델 자체 설계의 경량화를 통해 효율성을 개선
 - NMS(Non-Maximum Suppression)을 적용해도 여전히 많은 BBOX가 존재
 - ※ BBOX가 많아질수록 계산 비용 자체가 선형적으로 증가

- Key contribution

- Top-down HPE의 효율성은 샘플 다양성이 주된 원인이라는 점 최초 주장
 - 다양한 샘플에 적절한 연산 자원을 할당하여 효과적인 동적 프레임워크 제안
 - ※ Router와 두 개의 HPE(small/large) 모델을 통해 동적 할당
 - ResNet-50&HRNet-W32 조합을 통해 HRNet-W32 대비 정확도 유지하며 50% 속도 향상
 - 다른 backbone/dataset에서도 재학습 없이 적용 가능

DynPose¹⁾

• Introduction

▪ Sample diversity

- 샘플의 다양성(난이도)은 두 가지 측면에서 발생

- Pose diversity

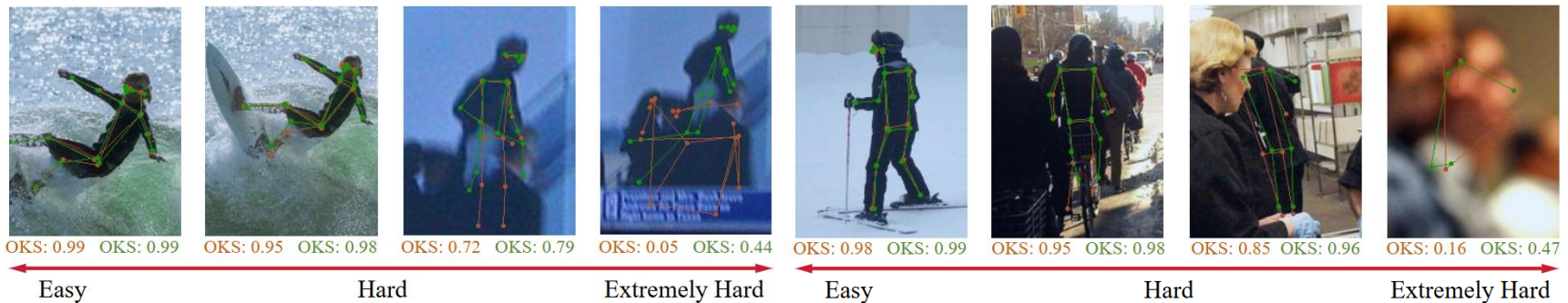
☼ 자세가 일반적인지 여부, occlusion 등 요인에 영향

- Box diversity

☼ 품질이 낮은 BBOX는 사람의 세부 정보를 놓치거나 불필요한 배경이 섞임

☼ 품질이 낮은 샘플의 경우 NMS를 적용해도 검출 모델이 많은 BBOX를 생성함

✓COCO val에서 6000개의 샘플에 대해 14만개의 BBOX를 만듦

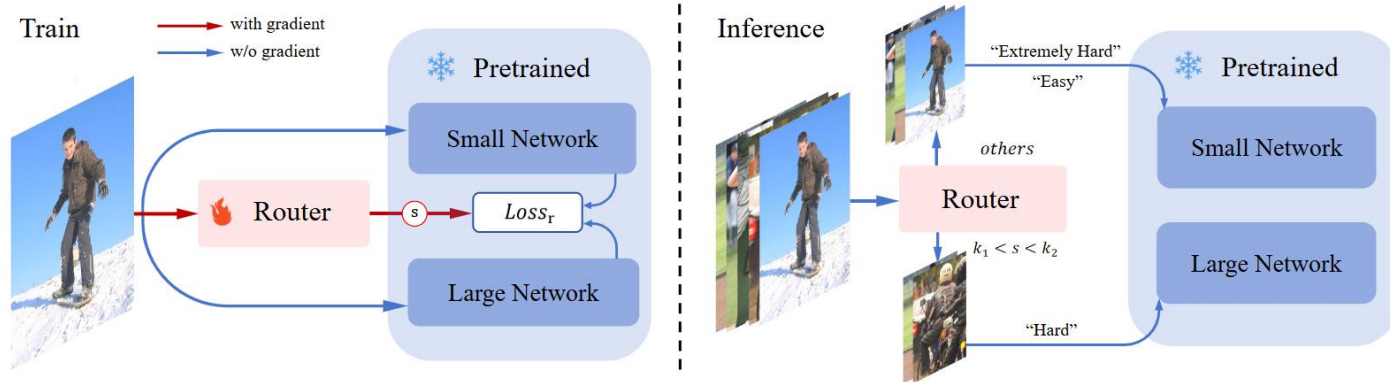


DynPose¹⁾

• Method

▪ Overall architecture

- 다양한 샘플에 맞춰 계산 자원을 동적으로 할당하여 추론 속도 향상
- 1개의 router와 2개의 HPE 모델(small/large)을 사용
- Router는 입력 샘플마다 두 모델 중 어디로 보낼지 결정
 - ⚡ Small network: 속도는 빠르지만 정확도 낮음
 - ⚡ Large network: 속도는 느리지만 정확도 높음
- 두 모델은 상호 교체 가능하며, 최종 결과를 생성할 때 사용



Overall architecture

DynPose¹⁾

• Method

▪ Overall architecture

- 샘플의 추정 난이도 점수 s

$$s = Router(I), s \in (0, 1)$$

※ 해당 점수는 pose diversity와 box diversity를 반영

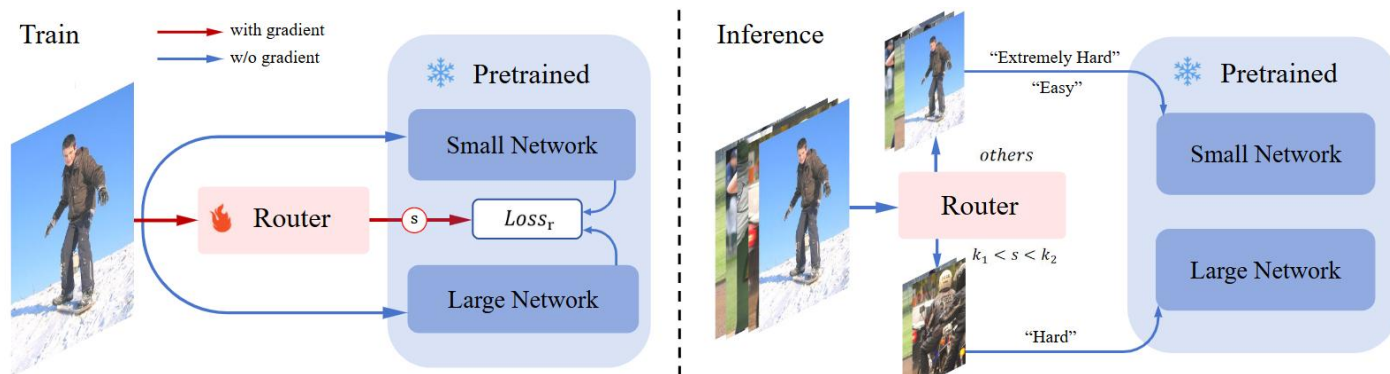
- 난이도 3단계

※ Easy: small과 large 두 모델 모두 정확한 예측 가능

※ Hard: small 모델에서는 어렵지만 large 모델에서는 정확하게 처리

※ Extremely Hard: 두 모델 모두 한계에 부딪힘

✓ 중복/오류 BBOX에서 유래하기 때문에 샘플 자체 의미가 없는 경우



Overall architecture

DynPose¹⁾

• Method

▪ Lightweight router

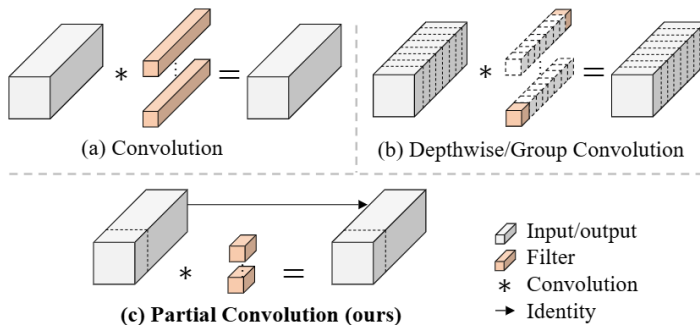
- Image $I \in \mathbb{R}^{3 \times H \times W} \rightarrow$ channel average pooling + max pooling $\rightarrow F_0 \in \mathbb{R}^{1 \times \frac{H}{2} \times \frac{W}{2}}$

- F_0 을 4단계의 convolution 기반 layer로 처리

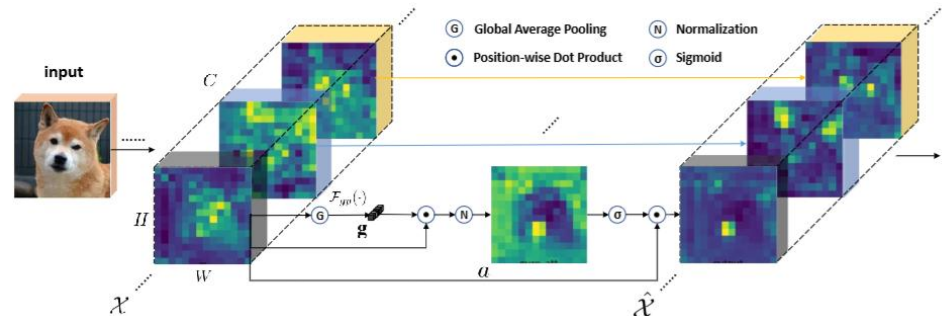
$$F_i = \begin{cases} \text{SGE}(C(F_{i-1})), & i \in \{1, 3\}, \\ \text{PConv}(C(F_{i-1})), & i \in \{2, 4\} \end{cases}$$

- F_4 를 flatten하여 최종적으로 F 를 얻고 이를 난이도 점수 s 로 변환

$$s = \sigma(W_2(\delta(W_1 F + b_1)) + b_2)$$



Partial covolution



Spatial group-wise convolution

DynPose¹⁾

• Method

▪ Optimization strategy

- Router 학습 목표

✧ 각 샘플의 추정 난이도를 평가할 수 있도록 점수 s 를 올바르게 supervise

✧ 난이도 지표로 small, large 두 모델 간 성능 차이를 사용

- OKS의 차이만 사용할 경우 두 모델 간 격차를 과소평가

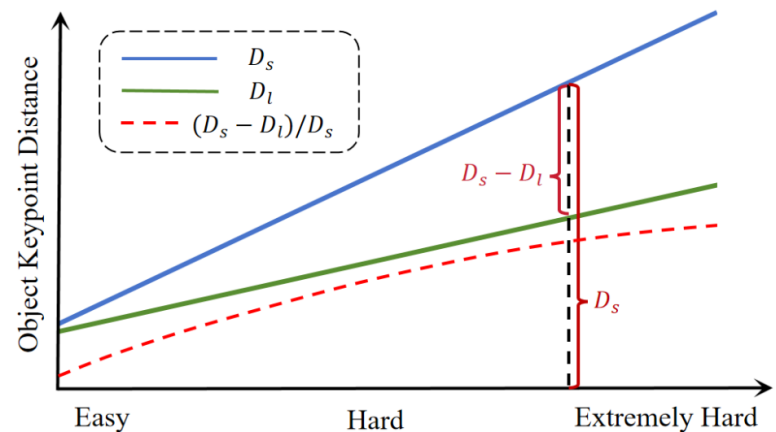
✧ 별도의 거리 척도가 필요

▪ OKD 제안

$$-D = \frac{\sum_{i=1}^{17} d^i \rho^i v^i}{\sum_{i=1}^{17} v^i}$$

▪ Router의 점수 s 에 대한 loss 정의

$$-\text{Loss}_r = \left(s - \frac{D_s - D_l}{D_s}\right)^2$$



OKD (Object Keypoint Distance)

DynPose¹⁾

• Method

▪ Inference strategy

- 추론 단계에서 각 샘플 난이도 s 에 따라 적절한 network 선택

- 난이도 점수 s 의 임계값 k_1, k_2 를 결정

※ 샘플 중 large network에 할당 비율을 조정하여 속도와 정확도의 trade-off 제어

- 테스트 샘플을 router를 거쳐 점수 s 를 얻고 오름차순으로 정렬

$$\text{※ } S_{\text{all}} = \{S_{\text{all}}^1, S_{\text{all}}^2, \dots, S_{\text{all}}^n\}$$

- 임계값 k_1 은 고정된 값 θ (hyperparameter) 사용

$$\text{※ } k_1 = S_{\text{all}}^1 + \theta(S_{\text{all}}^n - S_{\text{all}}^1)$$

- k_2 는 large 모델로 처리할 비율을 $p\%$ 에 맞춰서 설정

※ k_1 에 가장 가까운 S_{all}^j 파악

$$\text{※ } k_2 = S_{\text{all}}^{j + \lceil n \times p\% \rceil}$$

DynPose¹⁾

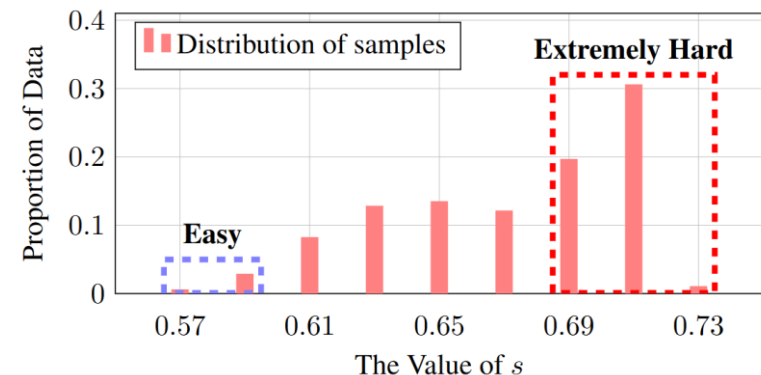
• Experiment

▪ Experimental setup

- MMPose 프레임워크에서 실험을 진행
- 데이터셋: MS COCO 2017
- 100,000개의 포즈 샘플로 학습, COCO val에서 성능을 평가
- Detection 모델은 Faster R-CNN을 사용

▪ Efficiency improvement of DynPose

Model	AP ₅₀	AP	FPS↑	FLOPs↓
ResNet-50	89.8	71.8	313	5.45G
Dyn-R50-H32/25	90.3	73.5	297	6.21G
Dyn-R50-H32/50	90.6	74.8	270	6.76G
Dyn-R50-H32/75	90.6	74.8	252	7.32G
HRNet-W32	90.6	74.9	223	7.69G
ResNet-101	90.4	72.8	261	9.01G
Dyn-R101-SL/25	90.5	74.0	151	17.19G
Dyn-R101-SL/50	90.6	74.3	106	25.18G
Dyn-R101-SL/75	90.6	74.3	81	33.16G
Swin-L	90.6	74.4	66	40.96G
ResNet-152	90.4	73.6	208	12.75G
Dyn-R152-VL/25	90.9	76.9	132	24.42G
Dyn-R152-VL/50	91.4	78.0	93	35.90G
Dyn-R152-VL/75	91.4	78.0	73	47.39G
ViTPose-L	91.4	78.2	59	58.68G



DynPose¹⁾

• Experiment

▪ Comparison with the SOTA

- 경량 backbone 기반 모델: TokenPose, PPT
- Distillation 기반 모델: OKDHP, DistillPose, SDPose
- SDPose는 경량 구조와 distillation 기법을 결합한 최신 SOTA 모델
- DynPose는 동일 조건에서 더 높은 AP와 효율 달성

Method	AP ₅₀	AP ₇₅	AP _M	AP _L	AR	GFLOPs	AP
TokenPose-S*[21]	88.7	79.0	68.3	78.5	77.0	4.7	71.8
TokenPose-B*[21]	89.5	80.2	70.1	79.8	78.7	5.2	73.2
PPT-S*[27]	87.7	76.8	66.1	76.7	75.1	2.0	69.8
PPT-B*[27]	89.5	80.8	70.3	79.8	78.8	4.7	73.4
OKDHP-2HG[23]	91.5	79.5	69.9	77.1	75.6	25.5	72.8
DistilPose-L[40]	89.9	81.4	71.0	81.8	79.8	10.3	†74.4
SDPose-S*[4]	89.5	80.4	70.1	80.3	78.7	4.7	‡73.5
Dyn-M2-H48(ours)	90.6	82.1	71.2	81.8	80.3	8.8	†75.0(↑0.6%)
Dyn-M2-H32(ours)	90.5	81.5	70.6	81.0	79.7	4.8	‡74.3(↑0.8%)

Compare with the SOTA

DynPose¹⁾

• Experiment

▪ Strong Generalization Ability

- 모델 간 일반화

☼ Dyn-R50-H32에서 학습한 router 파라미터를 다른 네트워크에 그대로 전이

☼ Small network: ResNet-50, MobileNet-V2, ShuffleNet-V2

☼ Large network: HRNet-W48, Swin-L, ViTPose-L

- 데이터셋 간 일반화

☼ COCO에서 학습한 DynR50-H32 router를 CrowdPose에 적용해도 성능 유지

Net_{small}	Net_{large}	FPS	FLOPs	AP
MobileNet V2	-	415	1.58G	64.8
ShuffleNet V2	-	457	1.37G	60.2
ResNet-50	-	313	5.45G	71.8
-	HRNet-W48	161	15.75G	75.6
MobileNet V2	HRNet-W48	222	8.86G	74.9
ShuffleNet V2	HRNet-W48	238	8.75G	75.1
ResNet-50	HRNet-W48	210	10.79G	75.4
-	Swin-L	66	40.96G	74.4
MobileNet V2	Swin-L	115	21.46G	73.9
ShuffleNet V2	Swin-L	123	21.36G	73.8
ResNet-50	Swin-L	108	23.40G	74.3
-	ViTPose-L	59	58.68G	78.2
MobileNet V2	ViTPose-L	110	30.32G	77.5
ShuffleNet V2	ViTPose-L	119	30.22G	77.6
ResNet-50	ViTPose-L	98	32.26G	78.0

Generalization across models

Model	AP ₅₀	AP	FPS	FLOPs
ResNet-50	80.8	63.7	313	5.45G
Dyn-R50-H32/25	81.8	65.6	297	6.21G
Dyn-R50-H32/50	82.3	66.9	270	6.76G
Dyn-R50-H32/60	82.4	67.3	258	6.98G
HRNet-W32	82.5	67.5	223	7.69G



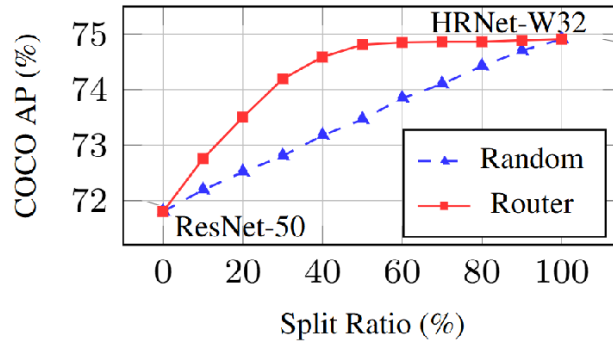
Generalization across datasets (CrowdPose)

DynPose¹⁾

• Experiment

▪ Ablation study

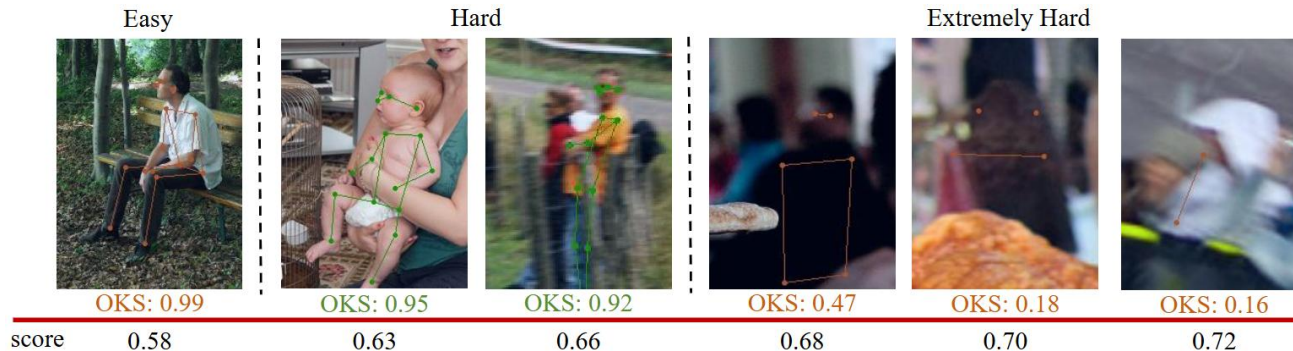
- Router 기반 추론 전략을 random 선택 전략과 비교하여 효율성 검증
- Router 경량화 설계를 위해 구성 요소 별 ablation 수행



Pooling	PConv	SGE	Time(ms)	AP
			0.52	74.64
✓			0.10	74.71
✓	✓		0.09	74.73
✓		✓	0.13	74.78
✓	✓	✓	0.10	74.81

Table 5. Ablation study of the router designs

▪ Visualization of results



DynPose¹⁾

- Conclusion

- 포즈 샘플의 다양성(난이도)가 top-down HPE 효율성의 핵심 요인임을 증명
 - 난이도별로 최적 모델에 샘플을 배정하는 DynPose 프레임워크 제안
- 동적 경로 선택
 - 추론 과정에서 router가 실시간으로 network 경로 결정
 - ※ Easy 및 Extremely Hard 샘플 → small network
 - ※ Hard 샘플 → large network
- 일반화 검증
 - 추가 실험에서 다양한 backbone, dataset에서도 성능과 효율성 유지
 - ※ 실무 적용 가능성 입증

ProbPose: A Probabilistic Approach to 2D Human Pose Estimation

[CVPR 2025]

ProbPose¹⁾

- Introduction

- Problem statement

- 기존 top-down HPE는 gaussian heatmap을 기반으로 keypoint 위치 표현으로 사용
 - ※ Gaussian heatmap을 실제 관절 분포와 동일하다고 보기 어려우므로 위치 오차 발생
 - Heatmap peak만 사용하고 분포 전체를 활용하지 않으므로 정보 손실 발생
 - 이미지 밖의 keypoint는 학습 및 평가에서 무시
 - ※ 학습 손실과 평가 지표 사이에 불일치가 발생하고 외부 keypoint 처리 성능이 낮음

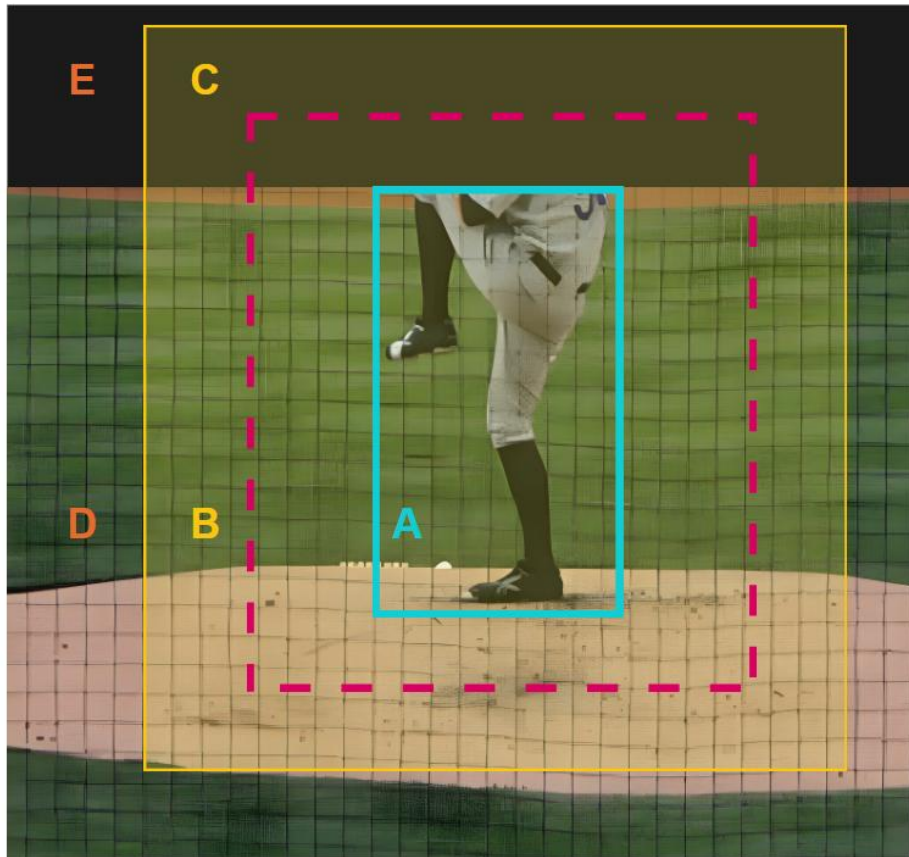
- Key contribution

- Probability map 도입: 각 픽셀에 keypoint 존재 확률을 직접 예측해 분포 활용
 - Expected OKSLoss: OKS 지표를 기대값 형태로 변환해 학습 손실에 사용
 - Presence probability: keypoint가 activation window 밖에 있을 확률을 함께 예측
 - ※ 이미지 외부 keypoint 정확도 45 % 상승
 - Random cropping: 경계 밖 데이터셋을 증강해 in-image/out-of-image 예측 모두 향상

ProbPose¹⁾

- Introduction

- On the types of keypoint



Area visualization

Dataset	A	B	C	D	E
COCO train	96.2	3.5	0.0	0.2	0.0
COCO val	95.8	3.9	0.0	0.2	0.0
CropCOCO val	68.8	2.2	23.5	0.1	5.3
OCHuman	99.2	0.8	0.0	0.0	0.0

Keypoint distribution

ProbPose¹⁾

• Method

▪ Probability maps and loss function

- 기존 방식에서는 고정된 gaussian heatmap 형태의 목표값을 따르며 MSE로 학습

- 학습은 빠르지만 출력 결과 자체가 gaussian 형태에 강하게 편향

※ 실제 keypoint 위치 분포가 gaussian 분포를 따른다는 이론적인 근거 부족

- ProbPose는 전통적인 heatmap 대신 probability map을 예측

※ Heatmap과는 다르게 Sparsemax activation을 통해 픽셀 합이 항상 1이 되도록 정규화

- Localization probability $p_L(x_i)$

$$p_L(x_i) = p(x_i \mid k_j \in AW, \text{img})$$

$$\sum_{x_i \in AW} p(x_i \mid k_j \in AW, \text{img}) = 1$$

※ 각 keypoint마다 개별 probability map을 가짐



ProbPose¹⁾

- Method

- Probability maps and loss function

- Expected Risk Minimization

- ※ OKS 기반 risk를 픽셀 별로 계산해 확률 분포에 가중

- Expected risk $R_{\text{exp}}(x_i)$

- ※ $R_{\text{exp}}(x_i) = (1 - \text{OKS}(x_i)) \cdot p_L(x_i)$

- 픽셀 x_i 에 대한 손실 $L_{\text{OKS}}(x_i)$

- ※ $L_{\text{OKS}}(x_i) = (1 - \alpha) R_{\text{exp}}(x_i) + \alpha g(x_i)$

- 예측 keypoint 좌표가 아니라 probability map 전체에 OKS를 적용

- ※ 낮은 OKS 픽셀엔 낮은 확률, 높은 OKS 픽셀에는 높은 확률을 부여

- $g(x_i)$ 는 Soble 연산자로 계산한 heatmap 기울기로 α 로 제어되는 정규화 항

- ※ 정규화 과정이 없으면 확률맵이 뾰족해져 과적합 위험

- Keypoint estimation 뿐만 아니라, 90% 확률로 keypoint를 포함하는 영역 같은 확률적 질문 가능

ProbPose¹⁾

- Method

- Probability maps and loss function

- 단순 argmax decoding

- ※ 확률맵에서 가장 높은 픽셀 하나를 선택

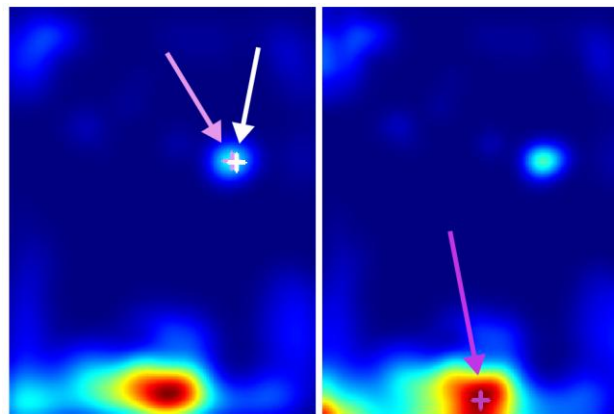
- ※ 해당 픽셀의 OKS도 얼마나 높은지는 고려하지 않음 → 분포 정보를 버림

- Expected OKS

- ※ $E_{\text{OKS}}(x) = \sum_{x_j \in \text{AW}} p_L(x_j) \text{OKS}(x_j; x)$

- ※ 전체 확률 분포를 고려하여 Expected OKS가 최대가 되는 위치를 선택

- ※ 단일 뾰족한 지점이 아니라 높은 확률이 넓게 퍼져 있는 영역을 선택



(a) Predicted probability map

(b) Expected OKS

ProbPose¹⁾

- Method

- Presence probability

- 기존 top-down HPE는 keypoint가 activation window 내/외부에 있는지를 명시적으로 다루지 않음
 - Activation window 밖인 경우 잘못된 예측이 발생해도 처리하기 어려움
 - Presence probability는 keypoint가 activation window 내에 존재할 확률을 의미
 - ※ BCE(Binary Cross-Entropy)를 기반으로 각 관절의 presence probability를 학습
 - ✓1은 내부, 0은 외부
 - Random cropping을 통해서 activation window 외부 사례를 생성하여 학습
 - 특정 임계값을 지정하여 임계값 미만 시에는 외부로 판단하여 별도 처리
 - ※ 단순 confidence 임계값보다 presence probability가 외부 판단에 더 효과적

ProbPose¹⁾

• Method

▪ Calibration

- 학습된 확률과 실제 발생 빈도가 어긋나면 신뢰도 해석과 임계값 선택이 불안정

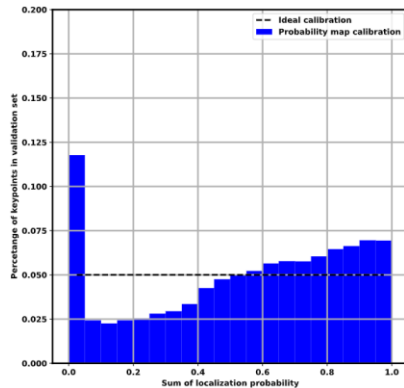
※ Probability map과 presence probability 모두 calibration을 진행

- Probability map은 상위 k%의 픽셀들이 전체 keypoint의 k%를 포함하는 방식으로 보정

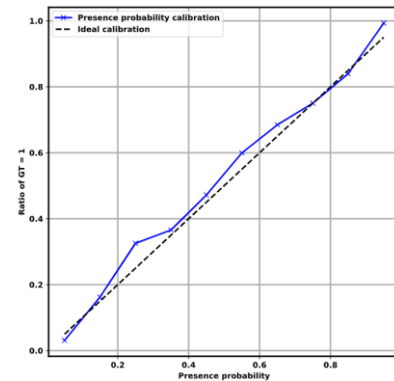
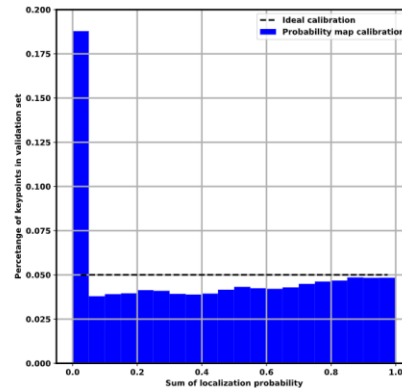
- Presence probability는 예측한 확률과 실제 발생 빈도를 일치시키기 위해 보정을 진행함

- CropCOCO dataset에서 temperature scaling을 적용해 이루어짐

$$z_i' = \frac{z_i}{T}$$



Probability map calibration curves



Presence probability calibration curves

ProbPose¹⁾

- Method

- Crop image augmentation

- 이미지 외부 keypoint 학습을 위해 별도 라벨링 없이 학습 데이터셋 구축

- ※ COCO 이미지를 무작위 crop, 이동, 스케일로 잘라서 activation window 구성

- ※ Crop strength(0~1) 값이 클 수록 더 잘라서 이미지 외부 사례를 많듸

- Probability map 학습

- ※ 원래 이미지 내에 있던 keypoint가 activation window 내에 남아 있지만 이미지 밖으로 밀림

- Presence probability 학습

- ※ 완전히 activation window를 벗어남



Crop image augmentation

ProbPose¹⁾

• Method

▪ Enlarging activation window size

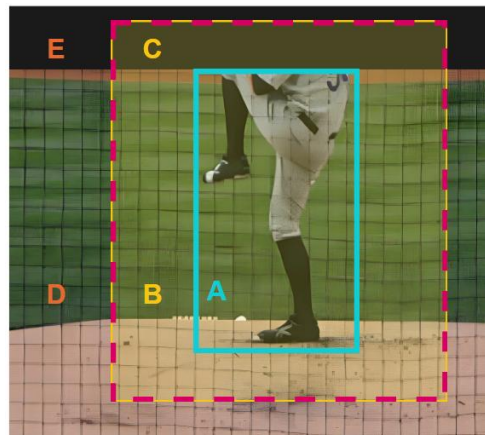
- 이미지 밖의 keypoint를 예측할 때 얼마나 멀리 떨어진 영역까지 볼지가 핵심

※ 단순히 activation window를 확장하면 일반적으로는 성능이 떨어짐

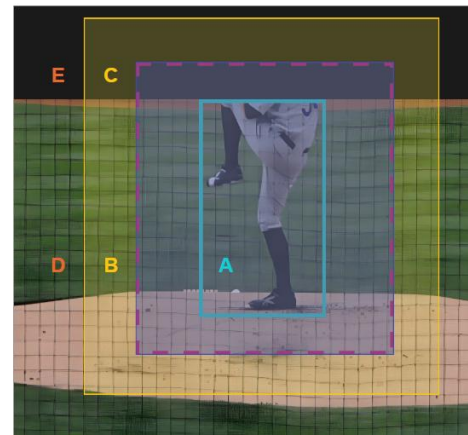
- Double heatmap 방식을 제안

※ Expert heatmap을 사용하여 모델 입력 크기와 일치하는 작은 activation window 내의 keypoint를 정밀하게 위치 추정

※ 더 큰 activation window에서는 넓은 probability map을 통해 멀리 있는 keypoint를 포착



(a) Traditional top-down method



(b) Double-heatmap approach

ProbPose¹⁾

- Data

- Evaluation metrics

- 기존 지표인 OKS와 PCK는 이미지 내부의 라벨링된 keypoint만 평가

- ※ 실제로 존재하지 않는 keypoint를 예측해도 계산에서 무시

- ProbPose는 keypoint가 activation window에 있는지를 판별하고, localization 함께 평가

- Activation window 밖의 keypoint를 평가하기 위해 Ex-OKS(Extended OKS)를 도입

- Ex-OKS

- ※ 예측 위치 x_i , presence probability를 평가함

$$\text{※ } d_i = \begin{cases} d_e(x_i^*, x_{i'}), & \text{if } p_p^* = 1 \wedge p_{p'} = 1, \\ d_e(AW, x_{i'}), & \text{if } p_p^* = 0 \wedge p_{p'} = 1, \\ d_e(x_i^*, AW), & \text{if } p_p^* = 1 \wedge p_{p'} = 0, \\ 0, & \text{otherwise} \end{cases}$$

$$\text{※ Ex-OKS}(i) = \exp\left(-\frac{d_i^2}{2k^2\sigma^2}\right)$$

ProbPose¹⁾

- Data

- CropCOCO

- 모든 실험은 COCO 데이터셋에서 학습하고 진행
 - 이미지 밖 keypoint 성능을 평가하기 위해 새로운 데이터셋인 CropCOCO를 도입
 - ☼ COCO 검증 이미지를 무작위로 crop하여 일부 keypoint가 activation window 외부에 위치하도록 구성
 - 모델의 다른 도메인에 대한 일반화 성능을 하기 위해 OCHuman 데이터셋에서도 평가



CropCOCO



OCHuman

ProbPose¹⁾

• Experiments

▪ Compare with the SOTA

- ProbPose-s는 베이스라인인 ViTPose-s 대비 localization 성능 향상

- 가장 큰 성능 향상은 CropCOCO에서 관찰

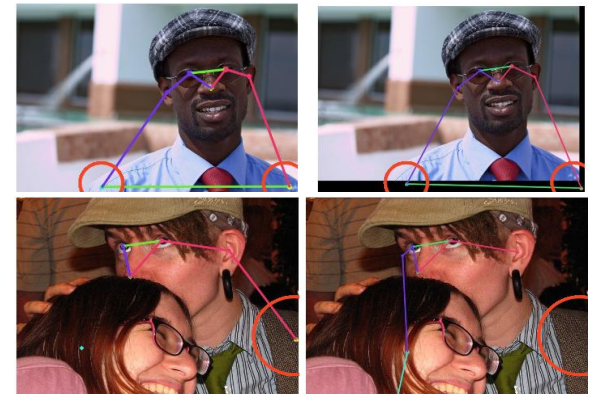
☼ ProbPose는 crop 증강을 통해서 이미지 밖 keypoint도 높은 성능을 달성함

- ProbPose-s-DH(double heatmap) 방식은 넓은 시야와 정밀도 사이의 trade-off로 인해 COCO와 CropCOCO에서는 성능이 소폭 하락

☼ 하락폭이 미미하며, OCHuman에서는 시야확장 효과로 인해 성능 개선

Model	COCO		CropCOCO		OCHuman	
	mAP	Ex-mAP	mAP	Ex-mAP	mAP	Ex-mAP
HRFormer-s [32]	75.2	74.6	70.9	64.3	60.3	60.0
PVTv2 [28]	72.0	71.5	70.7	63.1	58.5	58.1
SWIN-t [15]	73.5	72.9	71.3	65.0	58.1	57.9
ViTPose-s [31]	75.9	75.3	72.7	66.5	60.3	60.1
ProbPose-s	76.6	76.4	81.7	73.9	<u>60.4</u>	<u>60.2</u>
ProbPose-s-DH	<u>76.2</u>	<u>75.4</u>	<u>80.9</u>	<u>71.4</u>	61.4	61.2

Compare with the SOTA



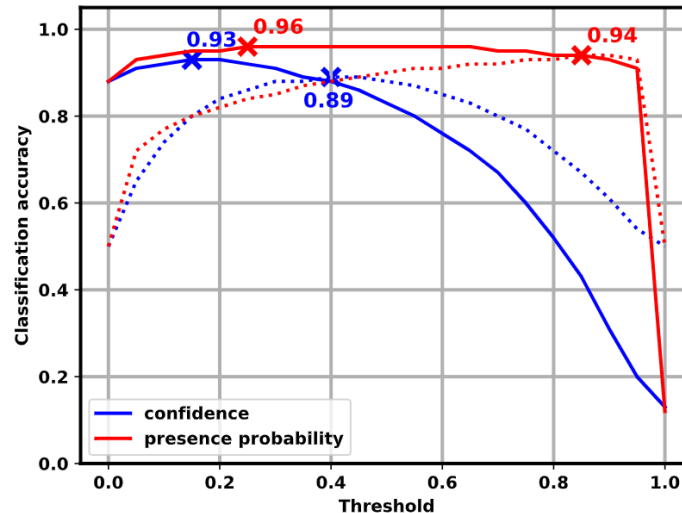
Labeling error

ProbPose¹⁾

• Experiments

▪ Presence probability vs confidence

- ProbPose는 keypoint가 activation window 내/외부에 있는지를 확인하기 위해 presence probability를 학습
- CropCOCO 전체 기준, presence probability가 분류 오류를 약 30% 감소
 - ※ Confidence는 위치 확률과 존재 판단이 섞여 있어서 보정이 어려움
 - ※ Presence probability는 존재만 따로 학습, 보정되어 안정적임
- Balanced subset(positive/negative 동일)에서는 오류율을 45%까지 감소시킴



ProbPose¹⁾

• Experiments

▪ Ablation study

- Probability map만 사용

⌘ COCO 데이터셋에서 keypoint localization 소폭 향상함

⌘ CropCOCO의 out of image에서는 성능이 크게 저하함

- Crop augmentation만 적용

⌘ COCO와 CropCOCO 모두 localization 성능이 향상하고, out of image 샘플 확보

- Probability map + crop augmentation

⌘ Crop augmentation만 적용했을 때와 동일한 성능 유지함

⌘ Probability map을 통해서 확률적인 해석이 가능해짐

crop	p-maps		COCO	CropCOCO
✗	✗		76.0	73.7
✗	✓		76.4	72.4
✓	✗		76.6	81.7
✓	✓		76.6	81.7

ProbPose¹⁾

• Conclusions

- Heatmap에서 probability map으로 전환하고, presence probability를 함께 예측하는 프레임워크 제안
- Expected OKSLoss로 학습 및 평가하고, Expected OKS 분포 전체를 활용하여 예측
 - 확률적인 해석이 가능
- Presence probability가 기존 confidence 임계값 기반 분류보다 activation window 내/외부 판정 오류를 크게 감소
- Crop augmentation으로 in-image/out-of-image 성능 모두 향상
 - CropCOCO 데이터셋 제안을 통해 out-of-image 성능 검증 가능
- COCO의 BBOX 경계 라벨링 문제는 완전히 해결되지 않음
 - 학습에서 완화되나, 평가 단계에서는 문제 발생

감사합니다