Learning Dense Hand Contact Estimation from Imbalanced Data

Daniel Sungho Jung[1], Kyoung Mu Lee[2]

[1]IPAI, [2] ECE & ASRI, Seoul National University

ComputerVisionLab Seoul National University

NEURAL INFORMATION PROCESSING SYSTEMS

arXiv

INTRODUCTION

Motivation

- **Hands** play a **central role** in **human interaction** and **human manipulation** with the physical world.
- **No existing method** for **hand-dedicated** contact estimation.
- **No existing body contact model** achieves **reliable accuracy** or **generalization** across diverse hand interactions.

➡ Developing a **dense hand contact estimation model** is essential.

Our Goal

- Estimate **vertex-level hand contact** on **MANO** hand model.
- Modeling **hand interaction** with the **world**.

* Figure from MOW dataset

Challenges

- All hand datasets have "imbalance" in **spatial location** (e.g., fingertip) of contact.

➡ **Spatial imbalance issue**.

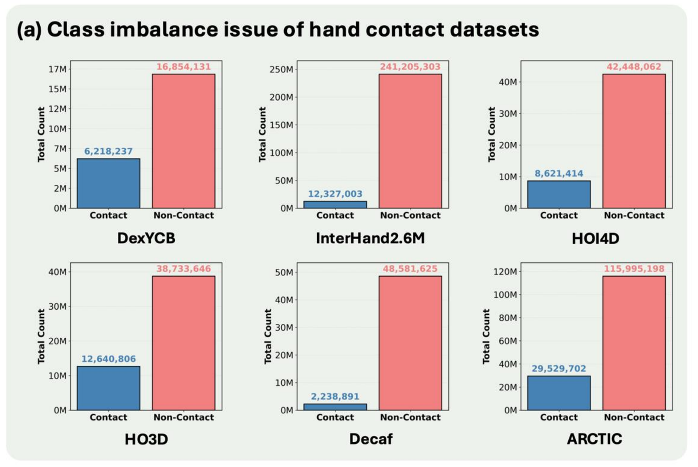
➡ Most of contact concentrated in **fingertip area**.

➡ **Difficult to generalize** on diverse contact patterns.

- All hand datasets have "imbalance" in **class** (contact vs non-contact).

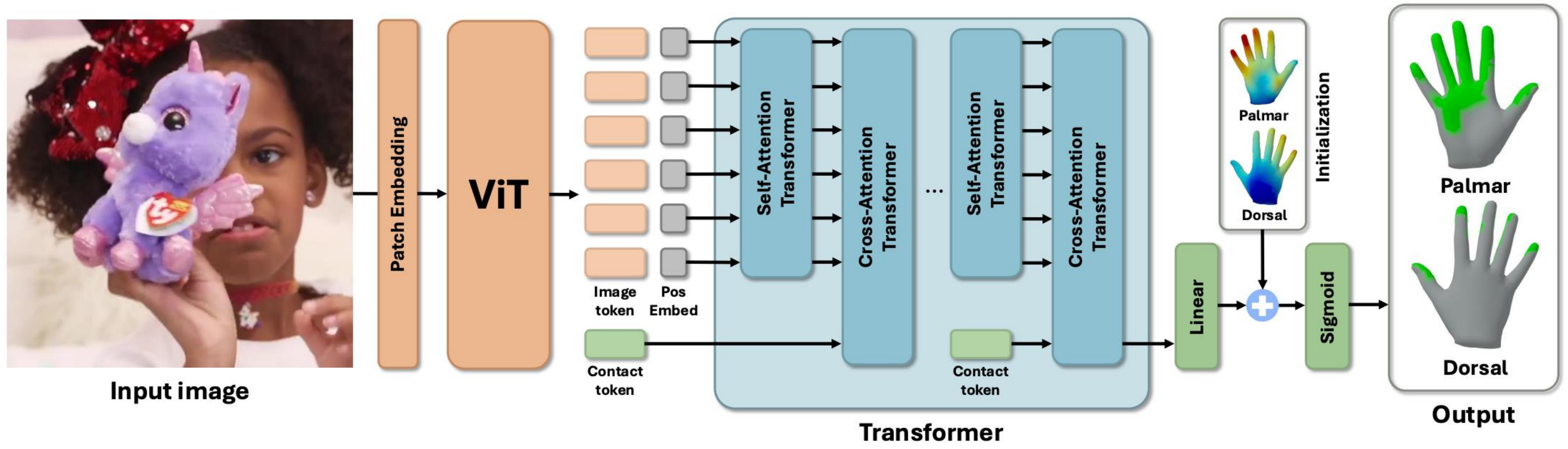
➡ **Class imbalance issue**.

➡ **Non-contact vertices** are **much more frequent** than contact vertices.



PROPOSED METHOD

Overall Pipeline of HACO



- **No bells and whistles**: Fully **Transformer-based architecture** trained in an **end-to-end manner**.
- **Contact initialization**: Learnable embedding for **effective initial contact estimation** based on **large-scale** hand datasets.

Vertex-Level Class-Balanced Loss

- Define the number of occurrence (n_y) of **class-balanced (CB) loss for each vertex** (v), which defines loss weighting factor (α).
- **Loss weighting factor** (α) is separately defined for **each vertex** (v), allowing **flexible loss contribution** to each vertex resolving **spatial imbalance issue** with vertex-level contact learning.

$$\alpha_y = \frac{1}{E_n^{(y)}} = \frac{1-\beta}{1-\beta^{n_y}}$$

$$\mathcal{L}_{\text{CB}} = \frac{1}{|V|}\sum_{v\in V}\alpha_y \ell_{\text{BCE}}(y_v, p_v)$$

➡

$$\alpha_{y_v} = \frac{1}{E_n^{(y_v)}} = \frac{1-\beta}{1-\beta^{n_{y_v}}}$$

$$\mathcal{L}_{\text{VCB}} = \frac{1}{|V|}\sum_{v\in V}\alpha_{y_v} \ell_{\text{BCE}}(y_v, p_v)$$

* $y \in \{0,1\}$ is ground-truth vertex contact.
* $E_n^{(y)}$ is the effective number of samples, which is the key component of CB loss.
* $\beta \in [0,1)$ is a hyperparameter that controls influence of sample count on weighting.

Balanced Contact Sampling

- We build **sampling bins** ($\mathrm{B} = \{B_1, B_2, \dots, B_k\}$) from existing hand contact dataset (D) with **contact balance score (s_i)** for each hand (H_i).
- **Contact balance score** (s_i) is defined to measures "whether a sample contains **more rarely contacted vertices ($\mathbf{c}_i^\top(\mathbf{1}-\bar{\mathbf{c}})$)** and **fewer frequently contacted vertices ($\mathbf{c}_i^\top\bar{\mathbf{c}}$)** relative to the dataset average ($\bar{c}$), which resolves **class imbalance issue**.

$$\mathcal{B}_k = \{\mathbf{H}_i \in \mathbf{D} \mid s_i \in [\tau_{k-1}, \tau_k)\}$$

$$s_i = \frac{1}{V}\left(\mathbf{c}_i^\top(1-\bar{\mathbf{c}}) - \mathbf{c}_i^\top\bar{\mathbf{c}}\right)$$

EXPERIMENTS

Ablation Studies

< Effectiveness of Vertex-Level Class-Balanced Loss >

Methods	Precision↑	Recall↑	F1-Score↑
CE loss	0.530	0.294	0.348
L1 loss	0.521	0.392	0.413
L2 loss	0.531	0.298	0.352
Focal loss [32]	0.518	0.387	0.409
CB loss [6]	0.484	0.534	0.465
CB Focal loss [6]	0.522	0.360	0.392
LDAM loss [3]	**0.532**	0.224	0.293
Asymmetric loss [52]	0.484	0.479	0.440
Poly loss [26]	0.528	0.324	0.371
VCB loss (Ours)	0.525	**0.607**	**0.522**

- **VCB Loss**: **Separate weighting factor** for each vertex based on **training data statistics**.
- **Resolves spatial imbalance issue** with vertex-level contact learning.

< Effectiveness of Balanced Contact Sampling >

Methods	Precision↑	Recall↑	F1-Score↑
w/o sampling	0.520	0.542	0.481
w/ sampling (Ours)	**0.525**	**0.607**	**0.522**

- **Higher Recall and F1-score** with balanced contact sampling.
- **More true positive predictions** on contact, showing **resolved class imbalance issue**.

< Effectiveness of Large-Scale Training>

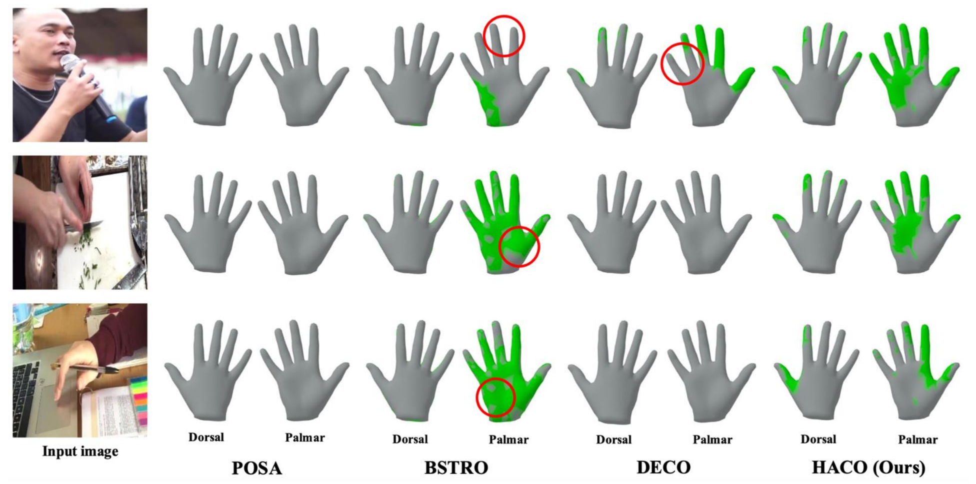
Methods	Precision↑	Recall↑	F1-Score↑
HACO trained on 1 dataset	0.498	0.348	0.373
HACO trained on 3 datasets	0.463	0.602	0.485
HACO trained on 14 datasets	**0.525**	**0.607**	**0.522**

- Large-scale training **consistently improves** performance.

Comparison to SOTA

< Dense hand contact estimation comparison with SOTA methods >

Methods	Precision↑	Recall↑	F1-Score↑
POSA [18]	0.134	0.128	0.101
BSTRO [22]	0.204	0.126	0.112
DECO [58]	0.246	0.235	0.197
HACO (Ours)	**0.525**	**0.607**	**0.522**

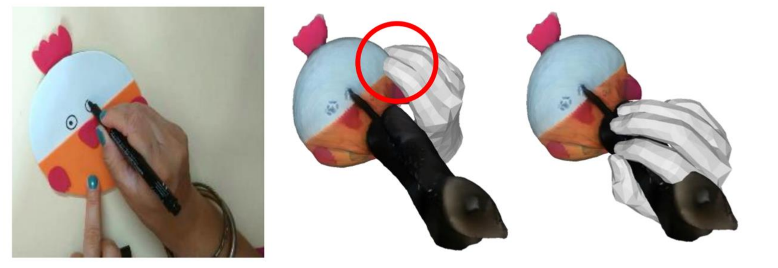


- HACO improves precision by 113%, recall by 158%, and F1-score by 164% compared to previous SOTA method.

< 3D hand and object reconstruction comparison with SOTA methods >

Methods	PVE↓	PJE↓	CD_{ho}↓	F-5_{ho}↑	F-10_{ho}↑
EasyHOI [35]	21.254	20.973	8.338	0.120	0.230
HACO (Ours)	**21.093**	**20.845**	**8.186**	**0.122**	**0.231**

- Replacing existing heuristic-based contact extraction pipeline to HACO leads to accurate 3D hand and object reconstruction.



CONCLUSION

- We propose **HACO** that addresses **class imbalance** and **spatial imbalance issue**, which are two challenges in learning dense hand contact estimation.
- Our HACO is **the first method for hand contact** trained on large-scale hand contact datasets, offering **generalization** to diverse hand contact patterns.