Hand Gesture Recognition in Buddhist Art Images: *Evaluation of a Keypoint-based Approach*

Li Weng

Zhejiang Financial College, China

Introduction

- Computer vision and machine learning have been used in:
  - Analyzing art works
  - Structuring heritage data
  - Assisting archeology
  - …

Art image analysis

- Related questions:
  - Style
  - Date
  - Role
  - Author
  - Object
  - Behavior
  - …

Hand gestures in Buddhist art

- Known as “Mudra”s in Sanskrit
- Each carries a particular meaning
- Hundreds of types in different branches

Varada

Dharmachakra

Anjali

Dhyana

Bhumyakramana

Abhaya

miscellaneous

miscellaneous

The objectives

- To understand the evolution of Buddhism in certain space and time
- To have a framework that can structure and link pieces of data
- To be able to carry out multiple analysis tasks

Related work

- Valentine Bernasconi, Eva Cetinic, and Leonardo Impett. 2023. A Computational Approach to Hand Pose Recognition in Early Modern Paintings

The challenges

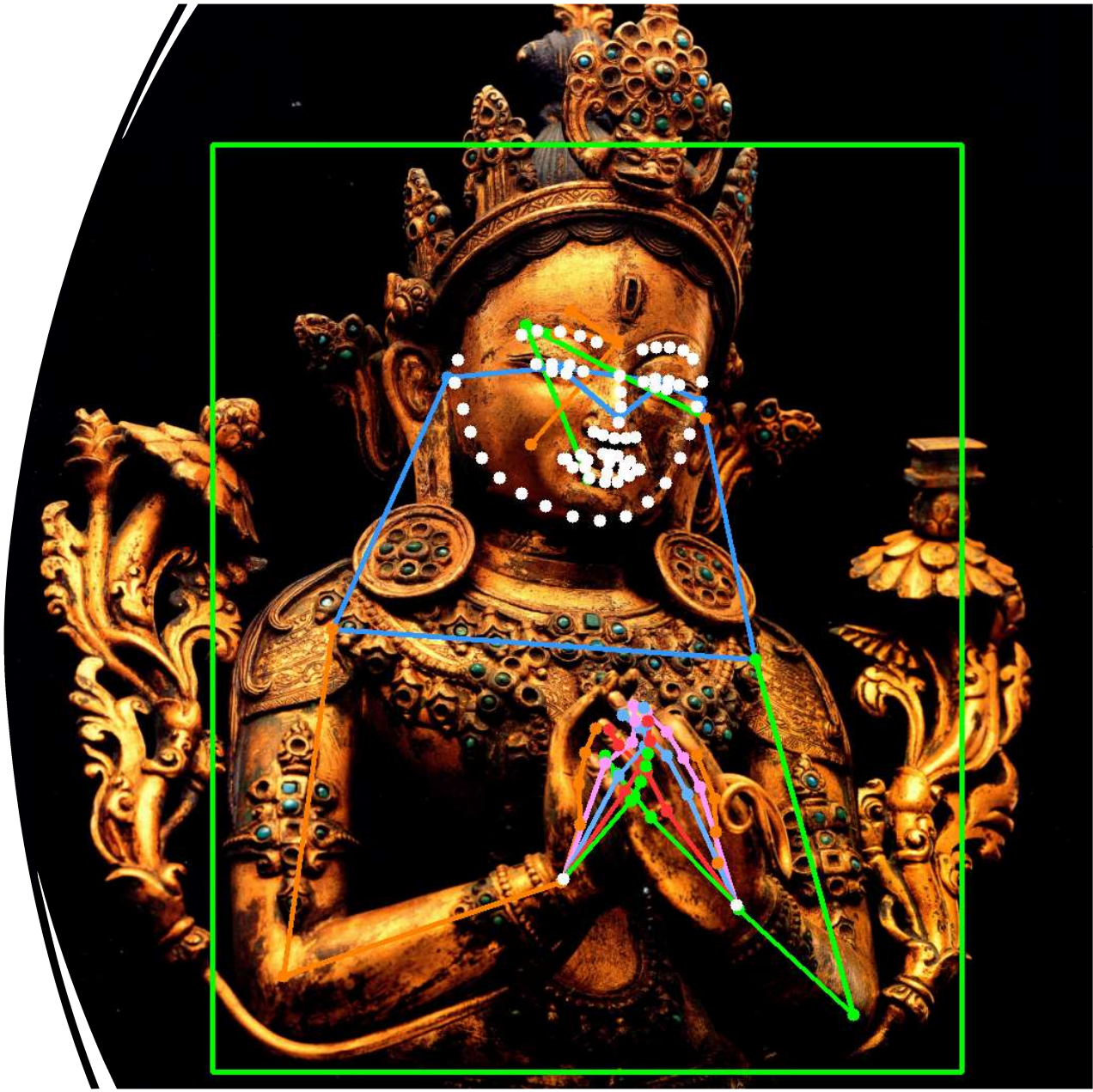
- Buddhist art does not always have a realistic style
- No pre-trained pose models exist

The proposed method

- A keypoint-based “detection and classification” approach
  1. Pose estimation
  2. Keypoint extraction and representation
  3. Classification of keypoint features

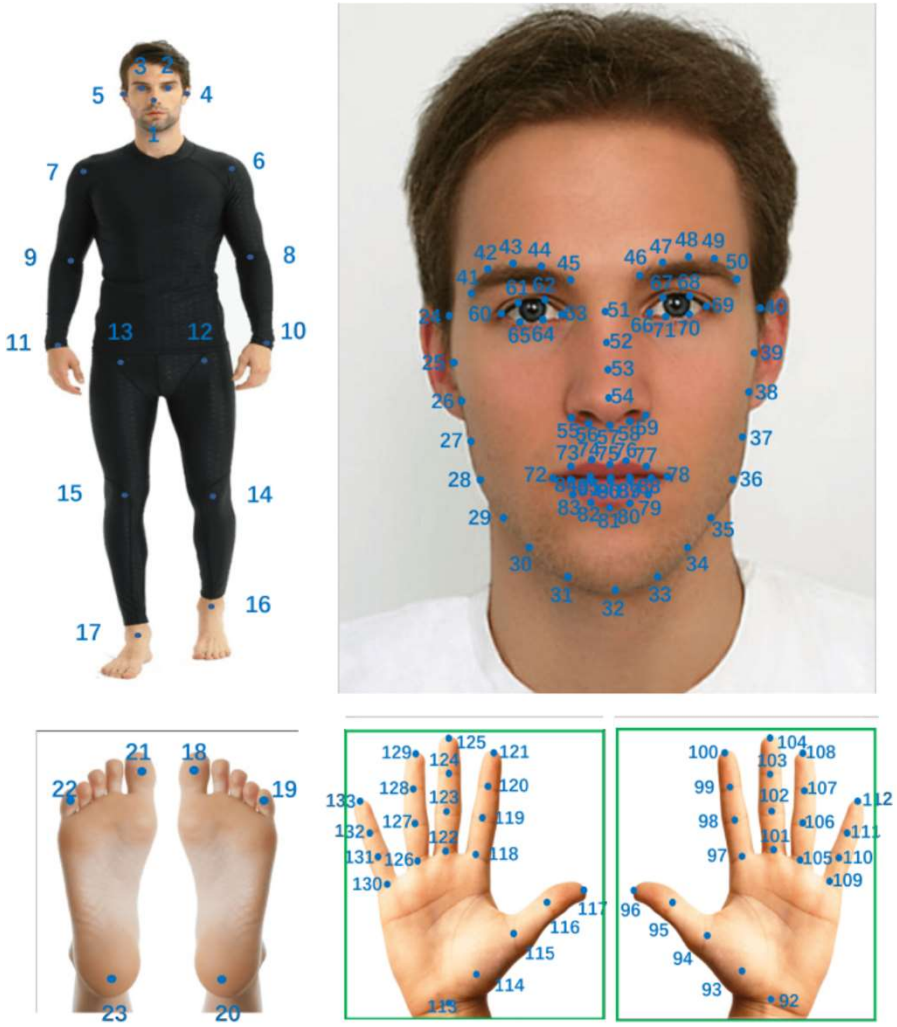
Keypoint-based pose estimation

- From bottom-up to top-down
- From low-resolution to high-resolution
- From multi-model to whole-body model



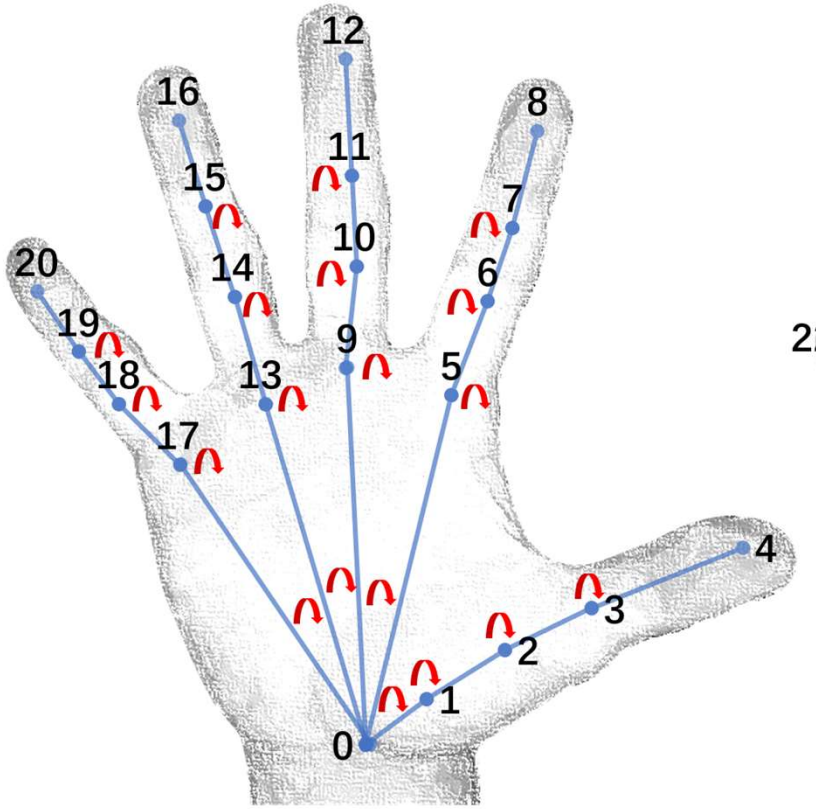
What's new in the paper

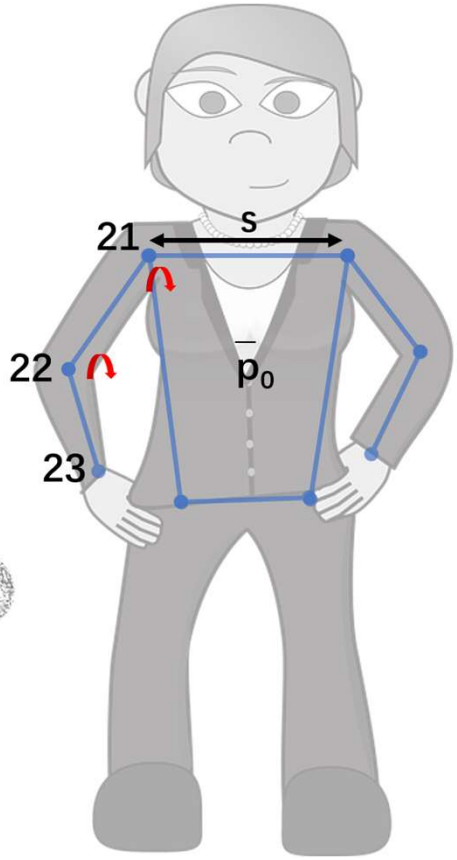
- Use of a high-resolution whole-body human pose model
  - Sheng *et al*. 2020. Whole-Body Human Pose Estimation in the Wild
- Revised hand feature representations



Hand feature representation

- Keypoint features (KP)
  - Angles and unit vectors
- Extended keypoint features (EKP)
  - Incorporate arms
- Normalized keypoint features (NKP)
  - Exploit spatial information





The dataset

category	count	simplified description
Dharmachakra	45	connected hands near chest
Abhaya	88	one hand up facing outside
Varada	80	one hand down facing outside
Bhumyakramana	60	one hand down facing inside
Dhyana	153	joined hands near belly
Anjali	38	joined hands near chest
miscellaneous	79	e.g. one hand holding something

Qualitative observations

- Statues are relatively easier to detect
- Faces are relatively easier to detect
- Human-like Buddhas are easier to detect

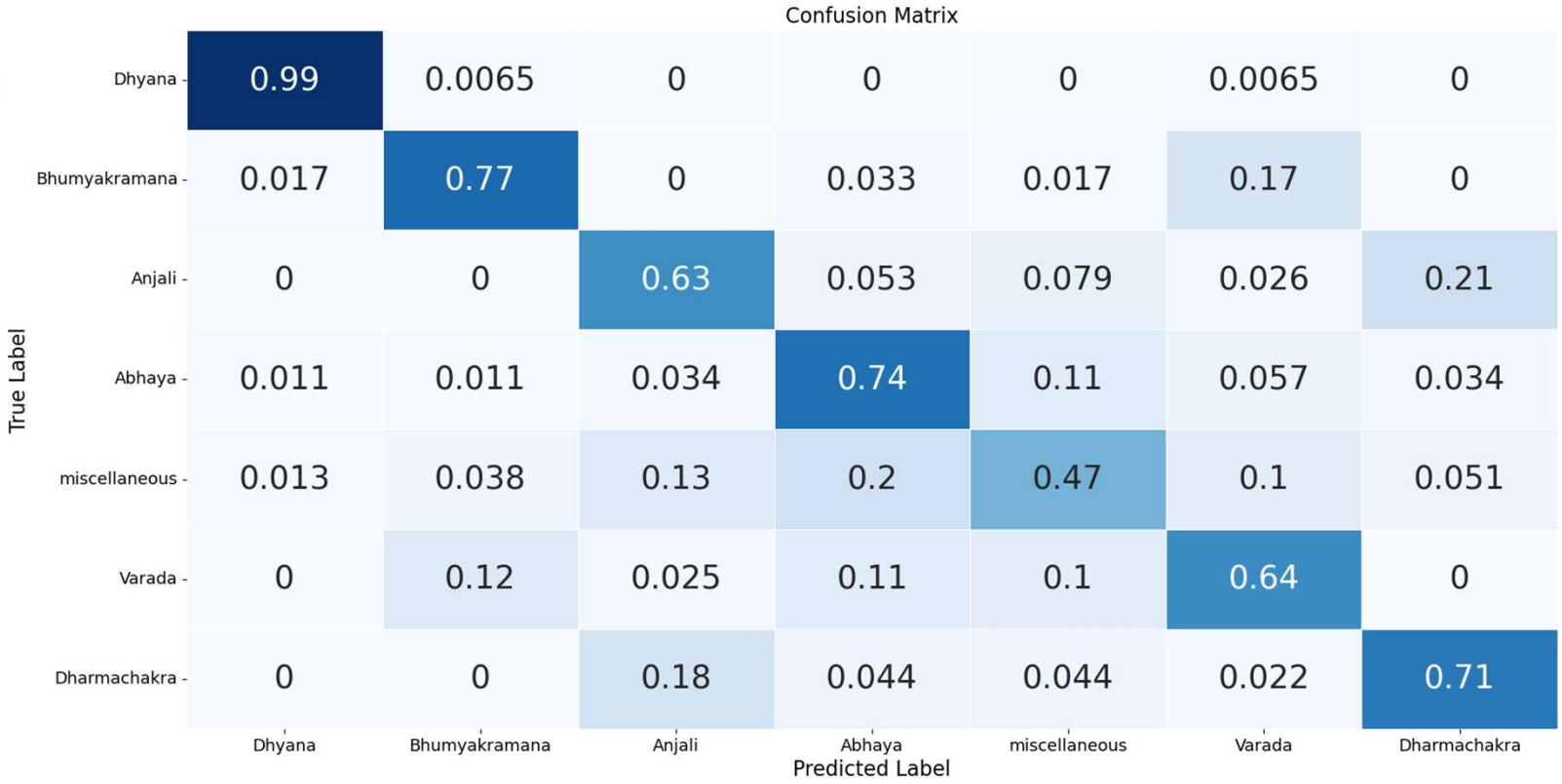
Classification performance

- Average accuracy > 0.7

classifier	accuracy score		
	KP features	EKP features	NKP features
SVM (linear)	0.64	0.67	**0.69**
SVM (RBF)	0.71	**0.72**	0.66
KNN ($k = 5$)	0.67	**0.68**	0.64
KNN ($k = 9$)	0.66	**0.67**	0.65
MLP ($h = 50$)	0.68	0.71	**0.74**
MLP ($h = 100$)	0.70	0.73	**0.75**

Classification performance

- Confusion matrix (MLP)



Conclusions

- Pre-trained whole-body human pose models are a good starting point
- Revised keypoint features prove to be useful

Future work

- Compare with other classification methods
- Integrate more features
- Fine-tune with style-transfer
- Distinguish each hand
- Gesture-based retrieval
- …

Q&A

THANK YOU FOR YOUR ATTENTION

LWENG@ZFC.EDU.CN