

ICCT 2026 · PAPER 1715



ThaiBuddhaPang



An Image Database and Benchmark for Classifying
Buddha Images by Pang

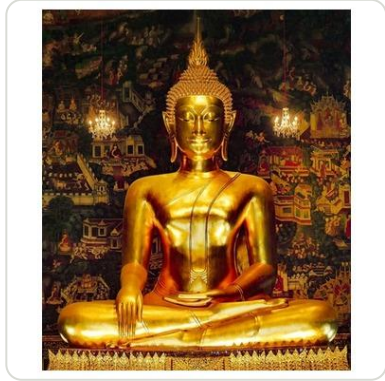
30 October 2026

Nattawat Duangsuwong · Jirathana Yosyingthummakul · Waree Kongprawechnon · Jessada Karnjana

Sirindhorn International Institute of Technology, Thammasat University · NECTEC, National Science and Technology Development Agency
Queen Mary University of London (QMUL), through the International Science Partnerships Fund (ISPF), administered by the British Council.

BACKGROUND

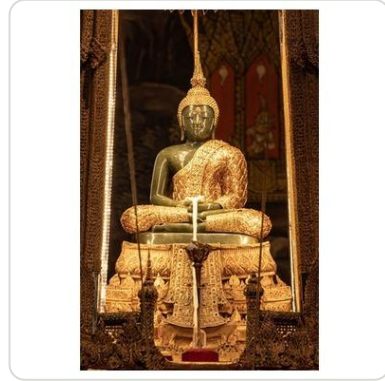
Pang: five iconographic poses of the Buddha



Marawichai

मारविच्च

Earth-touching: right hand over the knee, fingers pointing down



Samathi

สมาธิ

Meditation: both hands resting flat in the lap



Nakprok

นาคปรก

Naga-sheltered: seated under the hood of the naga



Prathanphon

ประทานพร

Boon-giving: one palm turned outward



Saiyat

ไสยาสน์

Reclining: lying on the right side

A pang is defined by posture, hand gesture, and symbols such as the naga hood, often a small part of the image.

OBJECTIVE

Research objectives

- 1 Build an openly licensed image dataset for classifying Buddha images by five pangs.
- 2 Benchmark four widely used pre-trained architectures on this dataset.
- 3 Find where and why the models make mistakes.

RESEARCH QUESTION

Which pre-trained architecture classifies pang best, and does cropping to the statue help?

PROBLEM

Why pang classification is hard

FINE-GRAINED

Classes look alike

Two pangs can share posture and hand position, and differ only by a hood above the head or the turn of a palm.

HIGH VARIATION

Each class varies widely

Style, material, age, lighting, viewpoint, and background all change within a single pang.

RESEARCH GAP

No dataset exists

To our knowledge, no public dataset for pang classification has been reported, and no benchmark of standard models.

THIS WORK

Two contributions

1

An open dataset

826 openly licensed photographs of 77 statue instances, with pang labels and statue bounding boxes.

The first public dataset for this task, as far as we know.

2

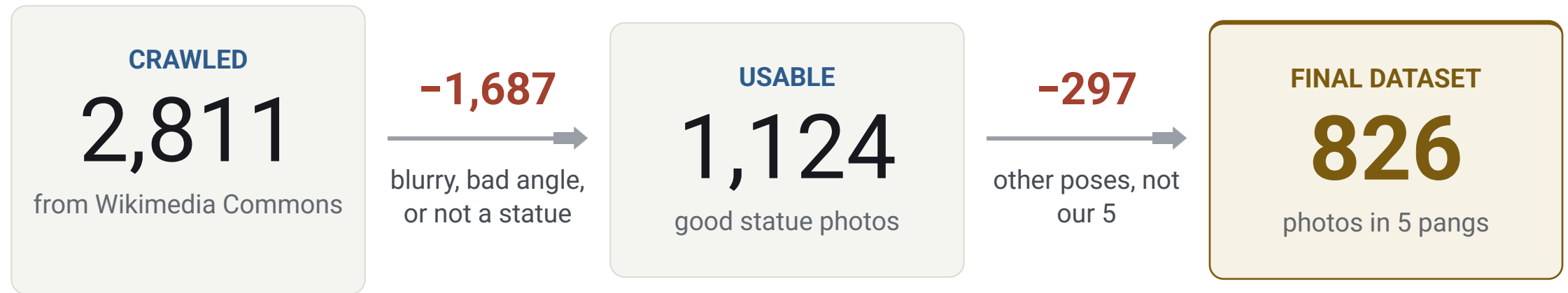
A benchmark

Four standard pre-trained architectures, trained and scored under one framework on the same folds.

Full and cropped inputs, scored with macro F1.

DATA COLLECTION

From 2,811 crawled images to 826



Then, for each photo :

1 Name the pose
which of the 5 pangs it is

2 Draw a box on the statue
to cut out the background

STATUE INSTANCES

What is a statue instance?

DEFINITION

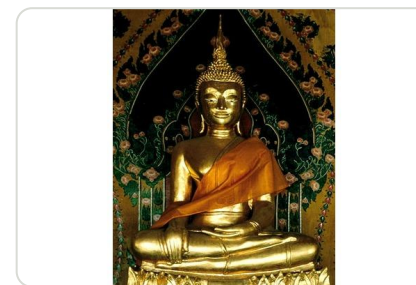
One physical statue, together with its replicas and close variants, and all of its photographs.

Different statues are different instances, even in the same pose (right).

A replica counts as the same instance as its original: the Emerald Buddha and its copies are one instance.

826 → 77 photographs come from only 77 statue instances

Same pose, four different statues: four instances



Wat Arun



Wat Pho



Wat Ratchabopit



Wat Benchamabophit

DATASET

826 photographs from Wikimedia Commons

826

photographs

5

pang classes

77

statue instances

All redistributable: CC BY-SA, CC BY, CC0, public domain. No NonCommercial or NoDerivatives images. Two annotators drew a pang label and a statue box for every image.

Pang	Pose	Images	Instances
Marawichai	Earth-touching	382	46
Saiyat	Reclining	146	11
Nakprok	Naga-sheltered	121	11
Samathi	Meditation	98	7
Prathanphon	Boon-giving	79	8
Total		826	77

Instance counts sum to 83 because five instances appear in more than one pang. The largest class is nearly five times the smallest.

THE LEAKAGE PROBLEM

With a random split, the model can recognise the statue instead of the pose.

77

distinct statues behind
all 826 photographs

71

photographs of a single
statue instance

45 / 79

Prathanphon images
come from one instance

SETUP

Models and evaluation protocol

Model	Family	Params
ResNet-18	Residual CNN	11.2 M
MobileNetV3-Large	Efficient CNN	4.2 M
EfficientNet-B0	Efficient CNN	4.0 M
ViT-Tiny	Transformer	5.5 M

ImageNet-pretrained, from *timm*. Head replaced with a five-way layer; all parameters fine-tuned.

EVALUATION

Five-fold cross-validation, class-stratified

INPUTS

Full photograph vs. cropped to the annotated statue box

METRIC

Macro F1, so every pang counts equally despite imbalance

NO HORIZONTAL FLIP

Which hand is raised is part of the label

RESULTS

ResNet-18 achieves the highest macro F1 on both inputs

Macro F1	Full	Cropped	Δ
ResNet-18	0.702	0.781	+0.080
EfficientNet-B0	0.617	0.708	+0.091
MobileNetV3-Large	0.634	0.704	+0.070
ViT-Tiny	0.624	0.613	-0.011
Mean	0.644	0.702	+0.057

BEST

0.781

macro F1, ResNet-18 on
cropped input

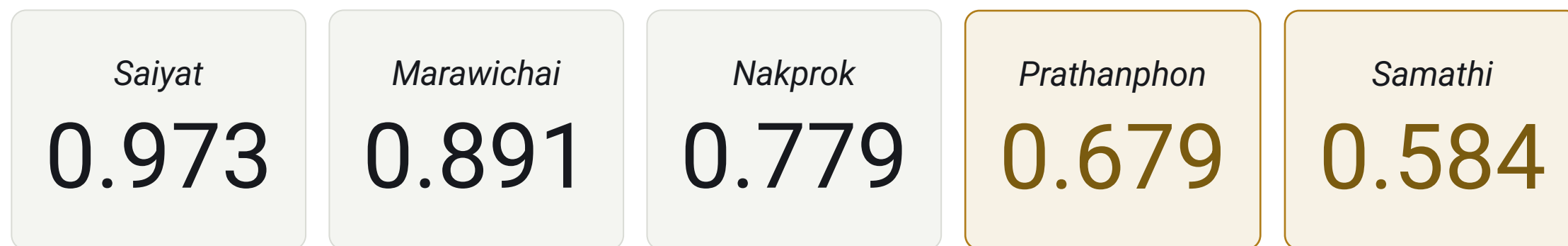
0.074 ahead of
EfficientNet-B0

Five-fold cross-validation. Same ranking in accuracy, precision, and recall.

PER-CLASS PERFORMANCE

F1 is highest for Saiyat and lowest for Samathi

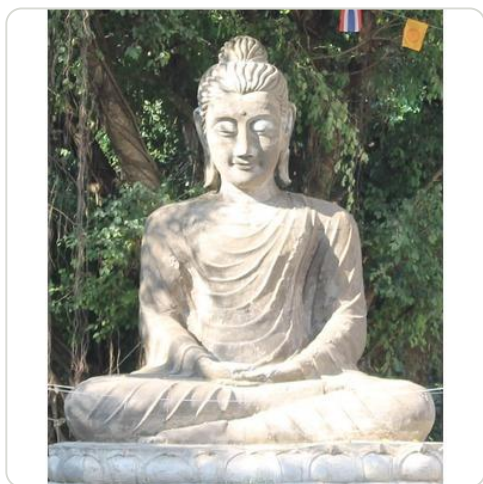
F1 per class · ResNet-18, cropped input



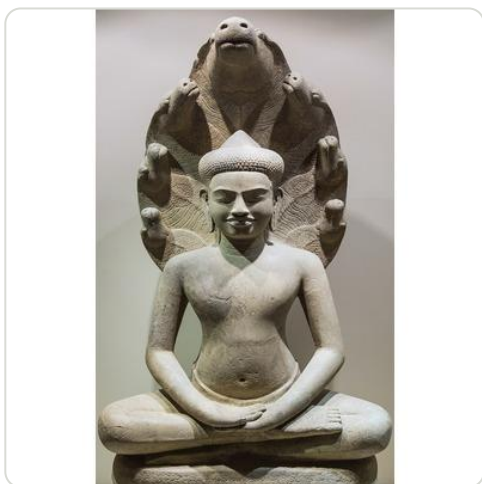
All 146 *Saiyat* images are classified correctly. The two lowest classes (gold) are analysed on the next two slides.

ERROR ANALYSIS

Samathi is often classified as Nakprok



(a) *Samathi*



(b) *Nakprok*

26 / 98

Samathi images predicted as *Nakprok*

Both have the same posture and hand position. The naga hood, the main difference, is above the head and can be cut off by a tight crop.

Most errors come from one statue instance: the Emerald Buddha and its replicas account for 46 *Samathi* images, and only 5 are classified correctly. Every other *Samathi* statue gets at least 2 in 3 images right.

ERROR ANALYSIS

Seated Prathanphon is often classified as Marawichai

15 / 20

seated *Prathanphon* images predicted as *Marawichai*

In both poses one hand rests at the knee. The difference is whether the palm faces outward or the fingers point down.

Standing *Prathanphon* is mostly right: 52 of 59 images. Seated: only 2 of 20.



Prathanphon, standing



Prathanphon, seated



Marawichai

SUMMARY

Conclusion

Dataset	826 openly licensed photographs of 77 statue instances, with pang labels and statue boxes.
Benchmark	ResNet-18 performs best at 0.781 macro F1 on cropped input. Cropping adds 0.057 on average and helps all three CNNs.
Errors	Most errors depend on small regions: the naga hood above the head and the direction of the palm.
Future work	Detect the hands and head region before classification, and add more distinct statues.

Thank you

Questions are welcome.



OPEN DATASET



github.com/TommyTheSecond/ThaiBuddhaPang

Nattawat Duangsuwong
Jirathana Yosyingthummakul

Supported by the Sirindhorn International Institute of Technology (SIIT), Thammasat University, and Queen Mary University of London (QMUL), through the International Science Partnerships Fund (ISPF), administered by the British Council.

Backup slides

A1 Why these five pangs?

A2 How are statue instances built?

A3 Why not group by photographer, date, or GPS?

A4 Could cropping be automatic?

A5 Why does ViT-Tiny come last?

A6 Training details

A7 Accuracy, precision, recall

A8 What goes wrong with the Emerald Buddha?

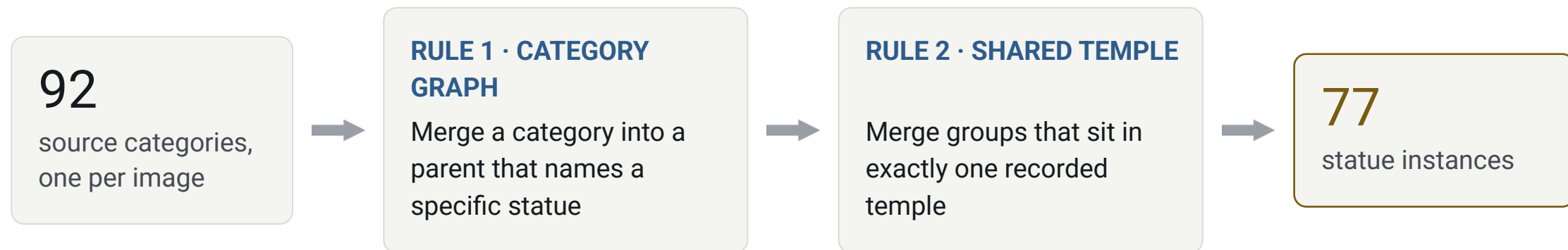
A9 Limitations

A10 Is the dataset openly usable?

Why these five pangs?

Frequency	Among the most frequently represented postures in Thai Buddhist art, per the Fine Arts Department reference [1].
Number of statues	Five-fold cross-validation over statues needs at least five per class. <i>Samathi</i> has 7 and <i>Prathanphon</i> 8, close to this minimum.
Scope	Other poses (for example abhaya, walking, alms bowl) were labelled <i>other</i> and excluded. Adding them is future work.

How are statue instances built?



Guard on rule 1

Nine categories naming a posture or region are never merged. The largest is the parent of 41 crawled categories.

Guard on rule 2

Groups spanning several temples are skipped. Without this, 182 images across four classes fused into one group.

Merging only

Categories are merged but never split, so an error can only make a group larger. This makes the evaluation stricter.

Union-find over categories. Temple recorded for 63% of images; six Thai/English temple spellings unified. No hand-curated replica list.

Why not group by photographer, date, or GPS?

Signal we tested	Why it was rejected
Photographer	Collapses the whole corpus to 11 groups
Capture session (camera + date, 84%)	Chains unrelated statues on museum and festival visits; 77 groups become 45
GPS coordinates (38%)	Less strict (72 groups vs. 43 on the same 282 images) and cannot link a replica to its original
Visual similarity	Finds 7 of 47 true merges and invents 66 false ones, because statues of the same pang naturally look similar
Hand-curated replica map	Changed macro F1 by +0.011 cropped and -0.032 full: the effect changes sign and is within noise; it also requires expert knowledge

The final method uses only the public category graph, the recorded temple, and two short lists: nine posture or region categories and six temple spellings.

Could cropping be automatic?

DETECTOR · YOLOV8N, OUT OF FOLD

0.675

mean IoU against human boxes

91%

of images get a detection

0.527

IoU on *Saiyat*, the worst class

DOWNSTREAM · RESNET-18 MACRO F1, ONE RUN

Human box



No crop



YOLO box



Bars on a 0–1 scale. All three arms trained together in one run; compare within it, not with Table V.

Why does ViT-Tiny come last?

Model	Params	Cropped F1
ResNet-18	11.2 M	0.781
EfficientNet-B0	4.0 M	0.708
MobileNetV3-Large	4.2 M	0.704
ViT-Tiny	5.5 M	0.613

Ranked by cropped macro F1.

Model size

The two smallest models finish second and third.

Dataset size

826 images may be too few for a transformer to learn attention patterns, while CNNs have a built-in spatial prior.

Cropping and learning rate

Cropping gives no clear change (−0.011), and its best learning rate is 30× lower than ResNet-18's.

Training details

Shared	Value
Optimizer	AdamW
Batch size	32 train, 64 inference
Epochs	30
Schedule	3-epoch warm-up, then cosine
Loss	Class-weighted cross-entropy, smoothing 0.05
Input	224 × 224, ImageNet normalization
Seed	42

Chosen in advance by an earlier search on the same grouped folds. Only learning rate mattered.

One override. EfficientNet-B0 updates BatchNorm. Frozen, two identical runs differed by 0.20.

Augmentation. Random resized crop and color jitter (RandAugment for EfficientNet-B0). No flip.

Model	LR	WD
ResNet-18	3e-3	1e-4
MobileNetV3-L	3e-3	5e-2
EfficientNet-B0	1e-3	5e-2
ViT-Tiny	1e-4	5e-2

Accuracy, precision, and recall

Cropped input	Accuracy	Precision	Recall
ResNet-18	0.840	0.809	0.773
EfficientNet-B0	0.776	0.731	0.711
MobileNetV3-Large	0.787	0.707	0.709
ViT-Tiny	0.723	0.608	0.632

Precision and recall are macro-averaged over the five classes.

Why macro, not micro?

Marawichai is 382 of 826 images (46%). A micro-averaged score would mostly measure that one class. Accuracy runs higher than macro F1 for the same reason.

What goes wrong with the Emerald Buddha?

Five source categories, merged into one statue instance

Commons category	Images
Emerald Buddha	17
Jade Buddha of Chiang Rai	11
Replicas of the Emerald Buddha	9
Replica of the Emerald Buddha at Wat Phra That Doi Suthep	9
Replicas of the Emerald Buddha at Wat Phra Kaeo, Chiang Rai	2
One instance (46 of them Samathi)	48

5 / 46

Samathi images classified correctly

“Jade Buddha of Chiang Rai” never says “Emerald”, yet the Commons category graph linked it on its own.

Likely cause: the statue sits small under tall, tiered canopies, in the region where a naga hood would be.

Limitations

One random seed	Every result is a single five-fold pass at seed 42.
Input sensitivity	Re-encoded crops vs. in-memory crops shift macro F1 by up to 0.03. The 0.057 cropping gain is larger, but not by a wide margin.
Not only Thai	84% of <i>Prathanphon</i> and 64% of <i>Nakprok</i> images come from outside Thailand or record no country.
Missed replicas	Four known same-statue links are absent from the Commons hierarchy; adding them by hand moved macro F1 less than run-to-run noise.
One label, two poses	<i>Prathanphon</i> mixes standing and seated statues that look very different.

Is the dataset openly usable?

Licence of each of the 826 images



CC BY-SA 541 **CC BY** 178 **CC0** 69 **Public domain** 37 **Free Art License** 1

Licences

No NonCommercial or NoDerivatives material.

Attribution

Source URL for 100% of images; photographer credit for 99%.

Contents

Images, pang labels, statue boxes, statue-instance ids, and the five folds.



DOWNLOAD THE DATASET

github.com/TommyTheSecond/ThaiBuddhaPang