Analyzing Finetuned Vision Models for Mixtec Codex Interpretation

Alexander R. Webber, Gabriel Ayoubi, Justin Witter, Zachary Sayers, Amy Wu, Elizabeth Thorner, Christan Grant
University of Florida Data Studio | ufdatastudio.com

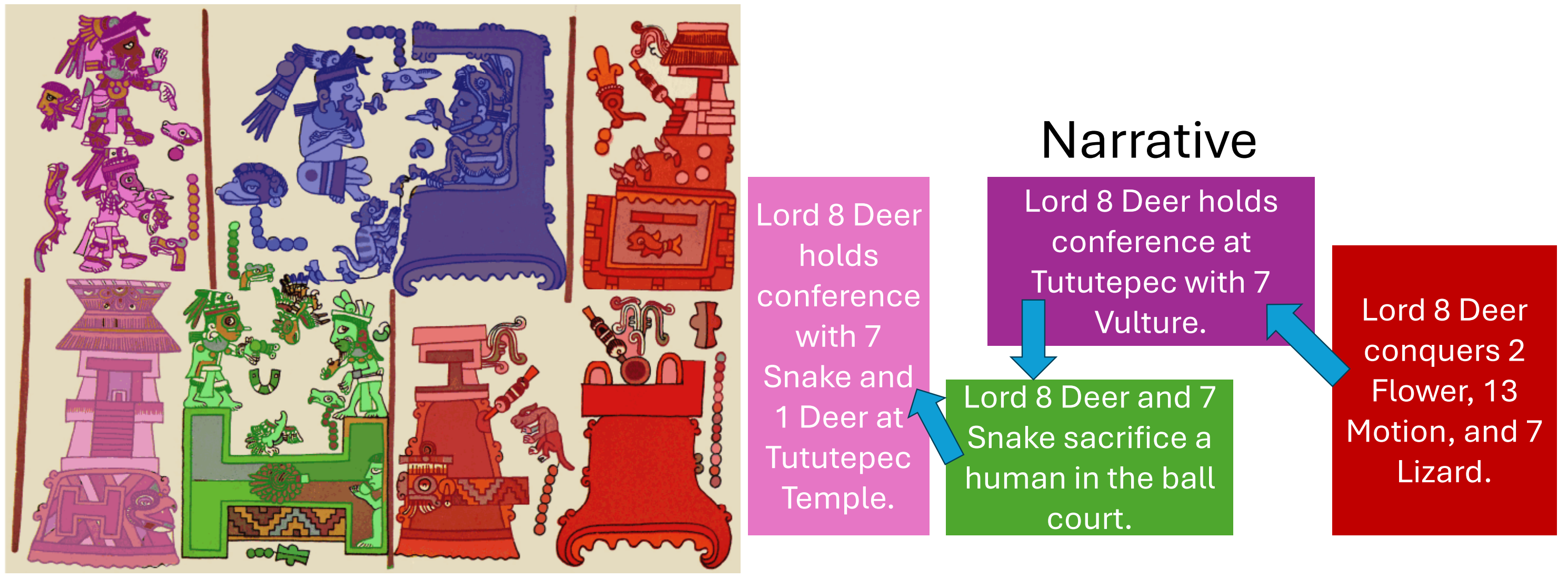
Abstract

The pre-Columbian Mixtec society recorded historical events through graphical media known as **codices**. These Mixtec codices are unique because the depicted scenes are **highly structured** within and across documents. As a first effort toward translation, we performed binary classification (**gender/pose**) on figures segmented from Mixtec codices. The results show that finetuned ViTs perform well on these tasks and are capable of identifying **similar conventions** to those outlined in Mixtec literature.

- We labeled a dataset of **1300 figures** extracted from three codices.
- We constructed binary classifiers for **gender** and **pose** by finetuning **VGG-16** and **ViT-16** on the novel dataset.
- We produced **attention maps** to visualize significant image sections during inference and **compare learned features** with expert opinions.

Mixtec Codices

The researchers labeled data from three popular sources: The Codices Vindobonensis Mexicanus I [4, 6], Selden [2, 1], and Zouche-Nuttall [5, 3]. Codices consist of scenes which are interpreted right-to-left in boustrophedon ordering. We rely on literature for interpreting these scenes and their constituent figures. Below is an example segmentation and narrative of four scenes from a codex page.



Narrative

Lord 8 Deer holds conference with 7 Snake and 1 Deer at Tututepec Temple.

Lord 8 Deer holds conference at Tututepec with 7 Vulture.

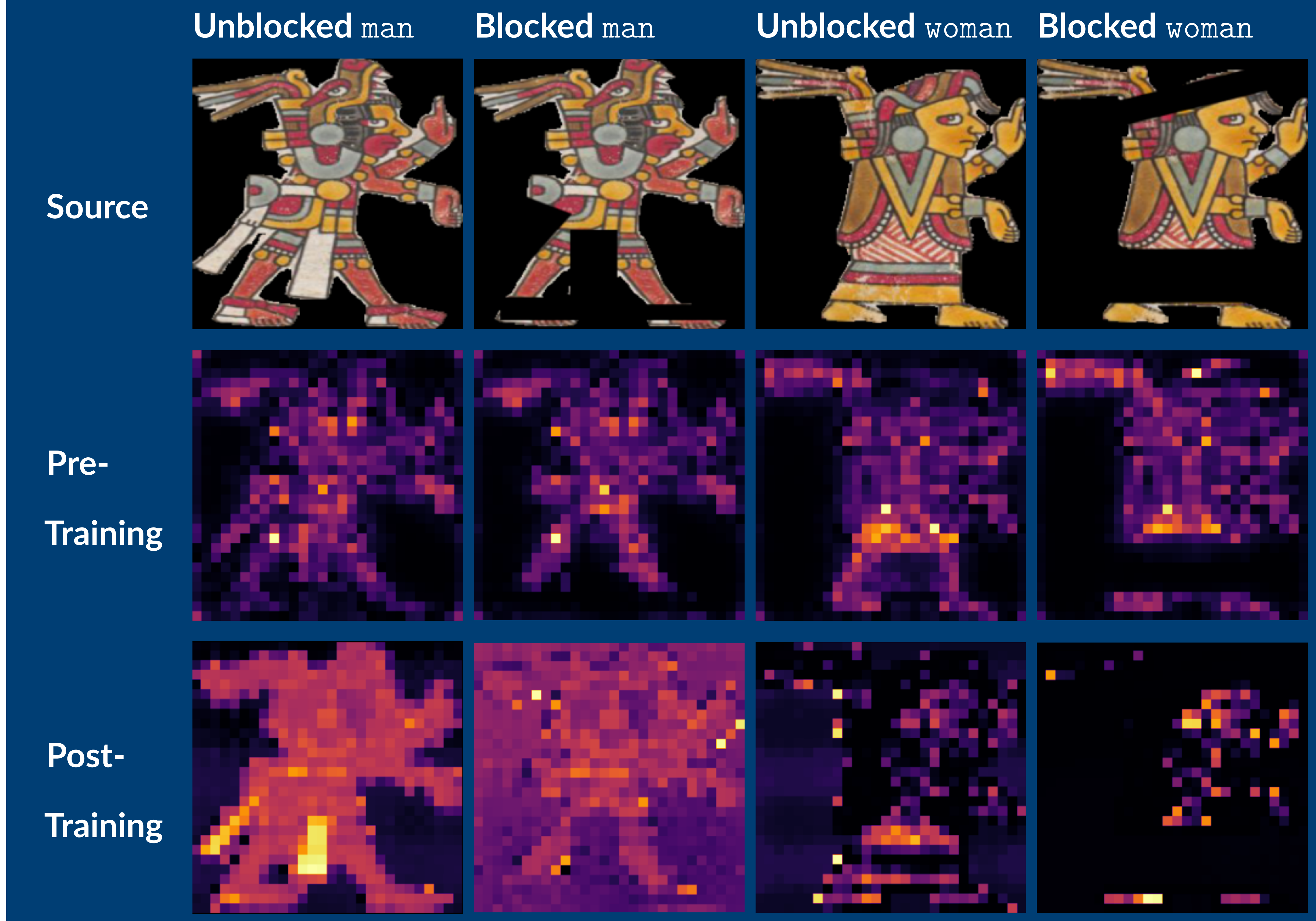
Lord 8 Deer and 7 Snake sacrifice a human in the ball court.

Lord 8 Deer conquers 2 Flower, 13 Motion, and 7 Lizard.

DATA STUDIO

Herbert Wertheim College of Engineering
Department of Computer & Information Science & Engineering
UNIVERSITY of FLORIDA

Finetuned vision transformers capture **semasiographic conventions** in Mixtec codices.



Use the QR Code to download the data

Dataset Statistics

Codex	Total Figures	Gender		Pose		Quality		
		Man	Woman	Standing	Not Standing	a	b	c
Nuttall	264	256	8	101	163	263	1	0
Selden	307	74	233	32	275	254	46	7
Vindobonensis	714	573	141	253	461	569	123	22
Totals	*1285*	*903*	*382*	*386*	*899*	*1086*	*170*	*29*

Reference Images

From left to right the image show a `man standing`, `woman not standing`, `woman standing`, `man standing`, `woman standing`, `man not standing`.

⇐ Attention Maps

ViT-16 Mean Attention Maps for `man` and `woman` show **increased attention** in the **loincloth area** for an unblocked `man`, and the **skirt area** for an unblocked `woman`, which follows expert opinion.

The blocked `man`'s weights do not converge to any particular area. The blocked `woman` did not produce meaningful activations.

Results

Model	A (%)	P (%)	R (%)	F_1 (%)
Gender	92.02 ± 1.52	92.55 ± 1.69	97.66 ± 1.78	95.02 ± 0.94
Pose	98.10 ± 1.24	97.77 ± 2.17	97.98 ± 1.65	97.86 ± 1.39
Orientation	96.24 ± 1.97	98.04 ± 1.45	95.47 ± 3.23	96.70 ± 1.78

The results show that VGG and ViT perform well when finetuned, with the transformer-based architecture (ViT) outperforming the CNN-based architecture (VGG) at higher learning rates.

References

[1] Liza Bakewell y Byron Hamann. *Codex Selden*. 2023. URL: http://www.mesolore.org/viewer/view/4/Codex-Selden.

[2] Alfonso Caso. "Códice Selden". En: *Sociedad Mexicana de Antropología, Mexico City* (1964).

[3] Sam Forstmann. *Codex Zouche-Nuttall*. 2023. URL: https://archive.org/details/codex-zouche-nuttall.

[4] Walter Lehmann y Ottokar Smital. "Codex Vindobonensis Mexic. 1". En: *Faksimileausgabe der mexikanischen Bilderhandschrift der Nationalbibliothek in Wien. Verlag von Anton Schroll & Co, Vienna* (1929).

[5] Zelia Maria Magdalena Nuttall. *Facsimile*. 1902.

[6] Unbekannt. *Bilderhandschrift: Sog. Codex mexicanus bzw. Codex Yuta Tnoho*. 1449. URL: http://www.onb.ac.at/sammlungen/hschrift/bibliographie.htm.