Supplementary Material
ArcAid: Analysis of Archaeological Artifacts using Drawings

1. Additional generation results on general images

Our algorithm works on examples it was not trained on and can be generalized for drawing generation of a variety of items and images. Examples on archaeological pots, an artistic relief, a sculpture and a roman coin are shown in Figure 10.

2. Additional generation results on CSSL

In addition to Figure 7, figures 11, 12 show examples of image-to-drawing generation on the CSSL dataset.

3. Hyper-parameters

As mentioned in the ablation study, the hyper-parameters are quite robust. For this experiment we used Resnet101[20], trained on 50% train and 50% test, 2-fold cross validation for shape classification. We used different values of γ_1 (similarity), γ_2 (Cross-Entropy) and γ_3 (generation). Table 9 shows that modifying γ values by ± 0.1, does not affect accuracy significantly.

4. Confusion matrices for classification tasks

We show the normalized Confusion matrices of all classification tasks, for 2 experiments in the cross validation setup, with and without our method, on top of Resnet101[20]. Figures 13, 14 and 15 show the confusion matrices for classification by shapes, 3 periods and 5 periods, respectively.

γ_1	γ_2	γ_3	Accuracy
0.8	0.05	0.15	**89.6%**
0.75	0.1	0.15	89.2%
0.7	0.15	0.15	88.3%
0.75	0.05	0.2	88.6%
0.7	0.05	0.25	89.3%
0.85	0.05	0.1	88.2%
0.9	0.05	0.05	89.0%

Table. 9. **Different values of γ.**

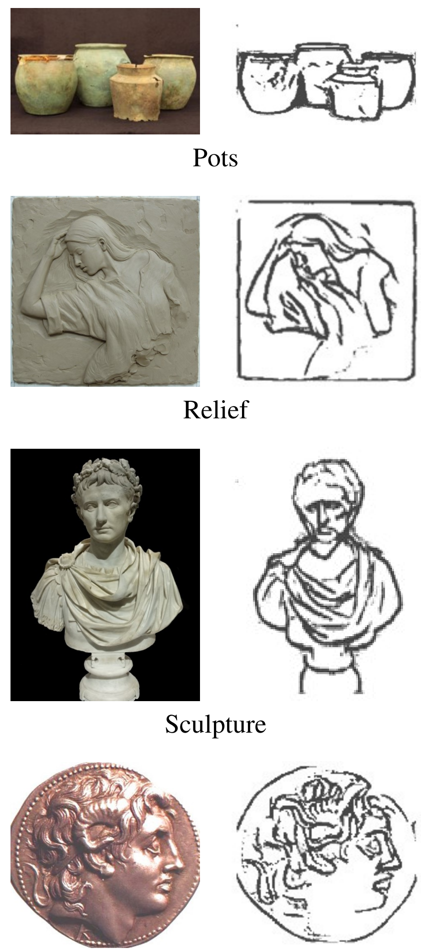
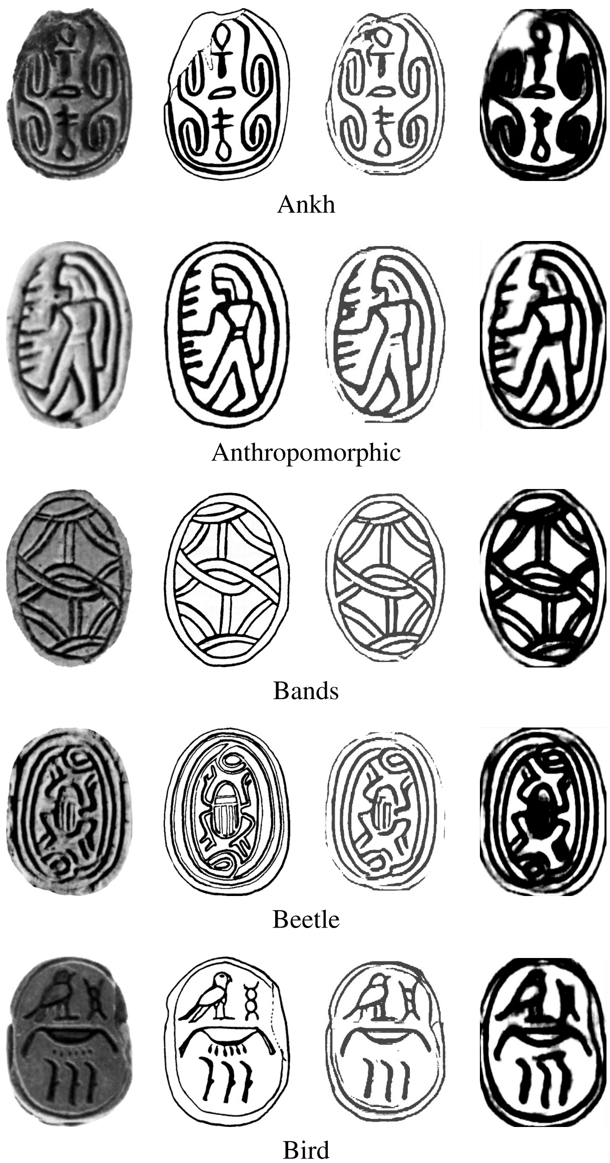


Fig. 10. **Our results on examples the model was not trained on.**

5. Classification and retrieval on 80/20 split

In addition to the 2-fold cross validation experiments on CSSL dataset, we also present the results obtained on a 80% train and 20% test split, 5-fold cross validation experiments. Table 10 shows the results of classification and retrieval by shape. Table 11 shows the results of classification by pe-



(a) Input (b) Manual (c) Ours (d) [43]

Fig. 11. **Image-to-Drawing generation: Ankh, Anthropomorphic, Bands, Beetle, Bird.**

riod. Our method improves all models. For classification and retrieval by shape, the best results are obtained when using our model on top of the DenseNet161 backbone. For periods classification, the best results are obtained with our method on top of both the DenseNet161 and CoinNet backbones.

6. Samples of classes from CSSL dataset

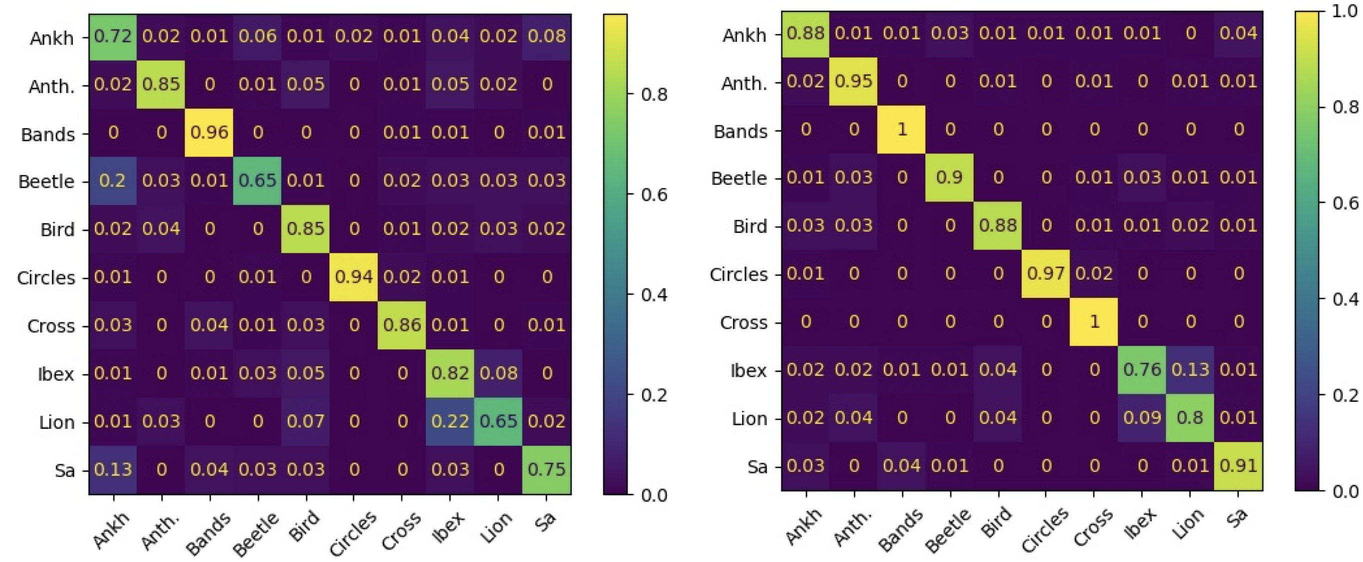
In addition to Figure 4, figures 16, 17 and 18 show different samples of all 10 classes of scarabs by shape from

(a) Input (b) Manual (c) Ours (d) [43]

Fig. 12. **Image-to-Drawing generation: Circles, Cross, Ibex, Lion, Sa.**

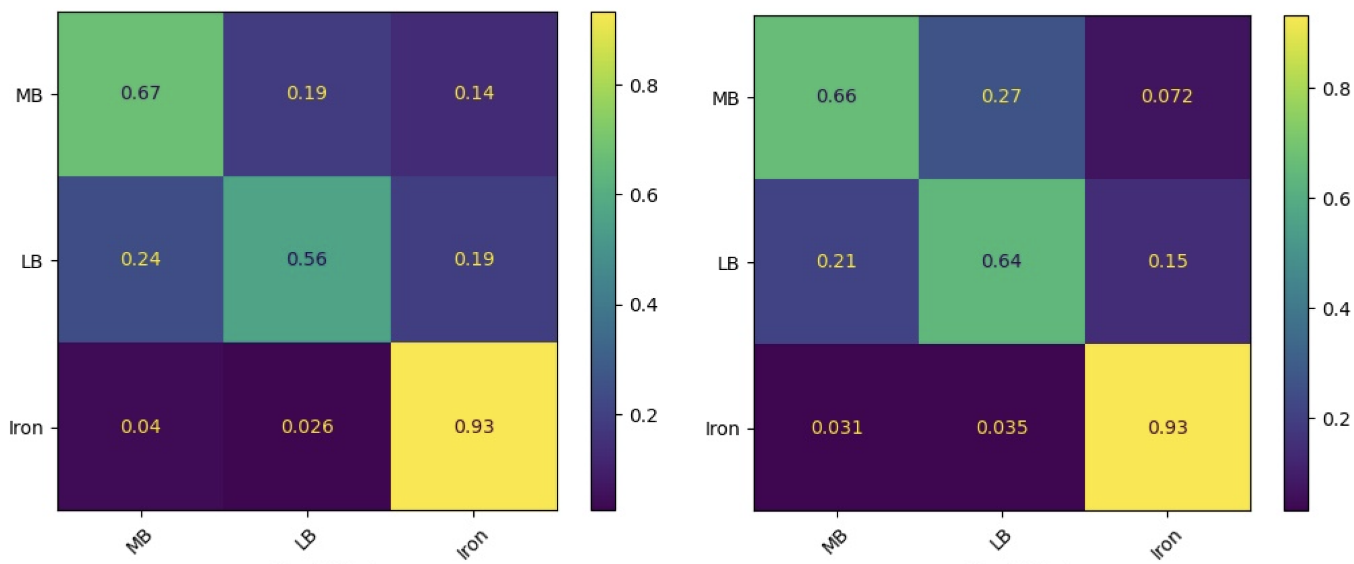
Model	Class.	P@1	P@10
DenseNet161[21]	88.6%	0.84	0.82
DenseNet161[21]+ours	**94.2%**	**0.92**	**0.91**
Resnet50[20]	85.9%	0.79	0.78
Resnet50[20]+ours	91.5%	0.88	0.87
EfficientNetB3[45]	80.7%	0.69	0.63
EfficientNetB3[47+ours	91.1%	0.89	0.87
Architectures for Archaeology			
CoinNet[2]	87.4%	0.82	0.81
CoinNet[2]+ours	92.8%	0.91	0.90
Glyphnet[6]	69.6%	0.56	0.52
Glyphnet[6]+ours	76.9%	0.67	0.64

Table. 10. **Shape classification and retrieval on our dataset for 80/20 split.**



(a) Only images (b) Images & drawings

Fig. 13. **Confusion matrices of classification by shape.**



(a) Only images (b) Images & drawings

Fig. 14. **Confusion matrices of classification by 3 periods.**

the CSSL dataset.

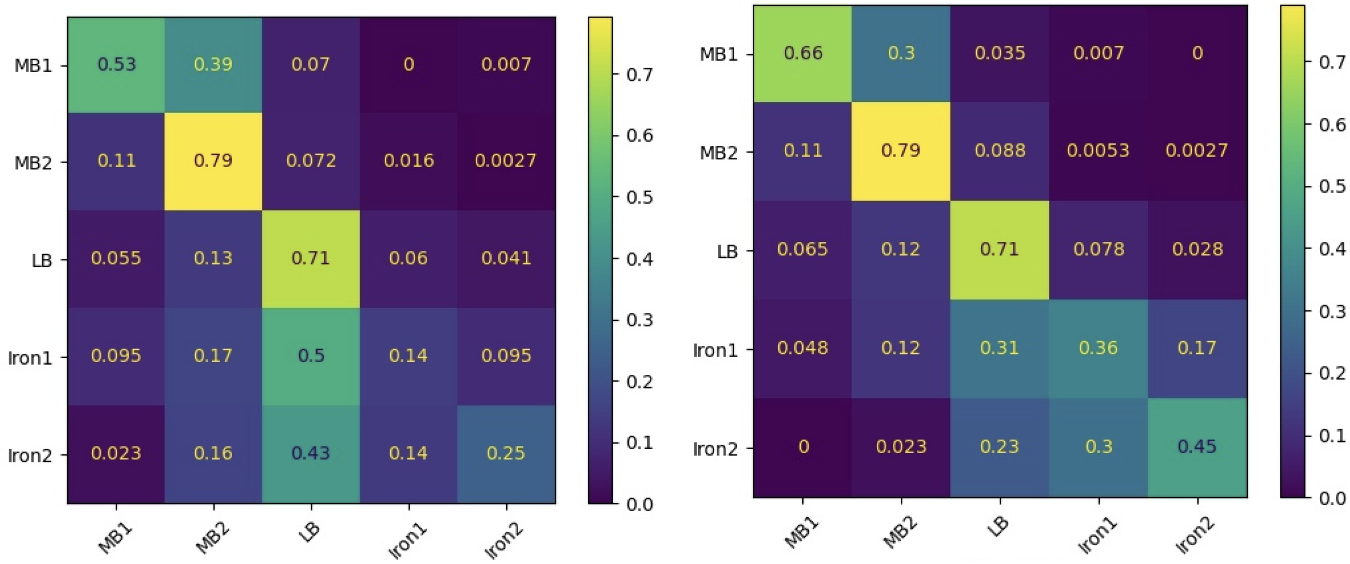
In addition to Figure 5, figures 19, 20 show multiple samples from all 5 sub-periods from the CSSL dataset.

7. Implementation details

Encoder/Decoder Architecture

For each backbone used as an encoder, we implemented a decoder, such that the encoder/decoder's architecture is similar to that of UNet [39]. We use skip connections between the encoder's layers and the decoder's up-sampling blocks, with unique implementations for Resnet50[20], DenseNet161[21], EfficientNetB3[45], Glyphnet[1,6] and CoinNet[2].

For image-to-drawing generation, $\tilde{D}$, the input for the generation loss, is the output of an NMS layer, fed by the decoder's output. The input for the similarity loss is the output of the encoder, a vector with different width for each



(a) Only images (b) Images & drawings

Fig. 15. **Confusion matrices of classification by 5 periods.**

Model	3 Periods	5 Periods
DenseNet161[21]	84.2%	72.1%
DenseNet161[21]+ours	84.6%	**73.0%**
Resnet50[20]	82.7%	71.5%
Resnet50[20]+ours	84.6%	71.9%
EfficientNetB3[45]	80.4%	68.1%
EfficientNetB3[45]+ours	84.8%	72.4%
Architectures for Archaeology		
CoinNet[2]	83.3%	72.6%
CoinNet[2]+ours	**85.4%**	**73.0%**
Glyphnet[6]	77.2%	62.9%
Glyphnet[6]+ours	77.5%	64.7%

Table. 11. **Period classification on our dataset for 80/20 split.**

encoder. The input for the Cross Entropy Loss is the output of the fully connected layer.

EfficientNetB3[45] was used in [38]. In [38], five copies of the backbone were trained, and then voting with them was used, demonstrating the results for classification and retrieval of archaeological artifacts. In our experiments, when we worked on our dataset, we trained and evaluated a single model, without voting.

Training parameters and details

The augmentations we use during training are random crop and horizontal flip. We also use each image and drawing as gray scale and transform them to a fixed size of 224/224. For each pair in a batch, we use the exact same augmentations. We use Resnet101[20], DenseNet161[21], EfficientNetB3[45] pre-trained on ImageNet, with learning rate of 0.00005. We train Glyphnet[1,6] and CoinNet[2] from scratch, with learning rate of 0.001 and 0.00005 respectively. For all models we train with batch size of 8 pairs (or photos). The optimizer we use is Adam, weight decay of 0.00001, and cosine annealing scheduler. The network we use for the perceptual loss is VGG16. We used NVIDIA RTX A6000 GPU for training.

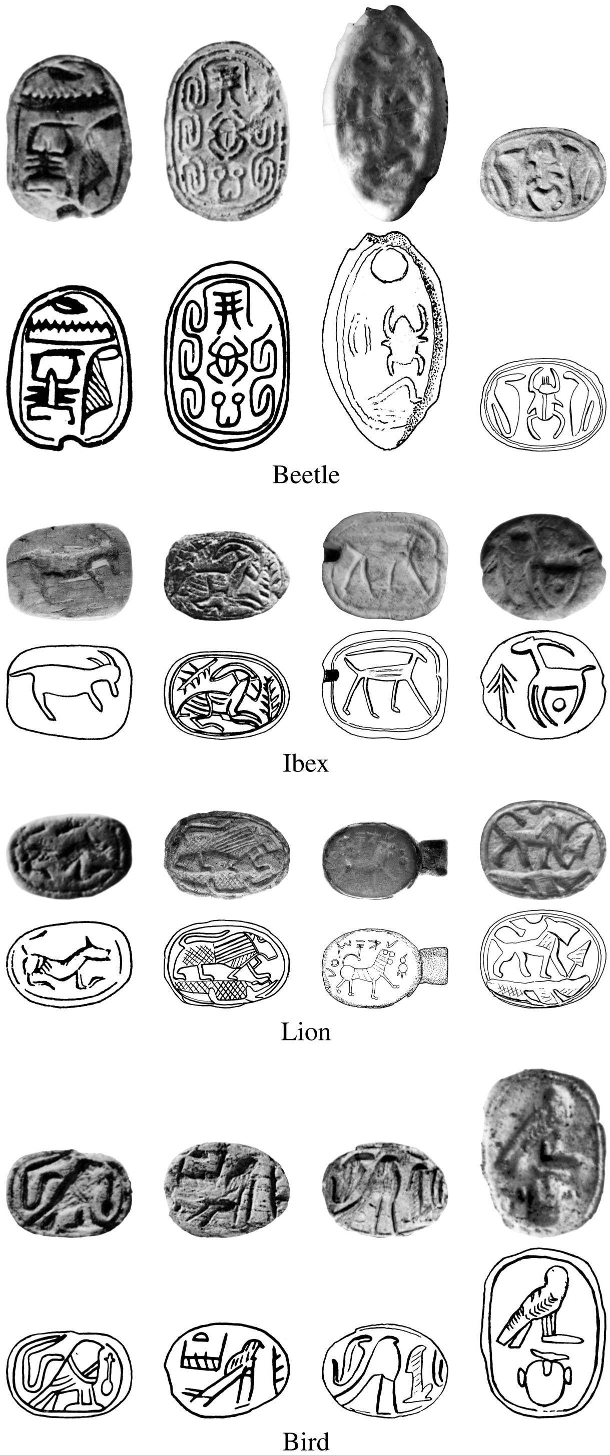


Fig. 16. **Samples of classes from CSSL dataset by shape: Beetle, Ibex, Lion and Bird.**

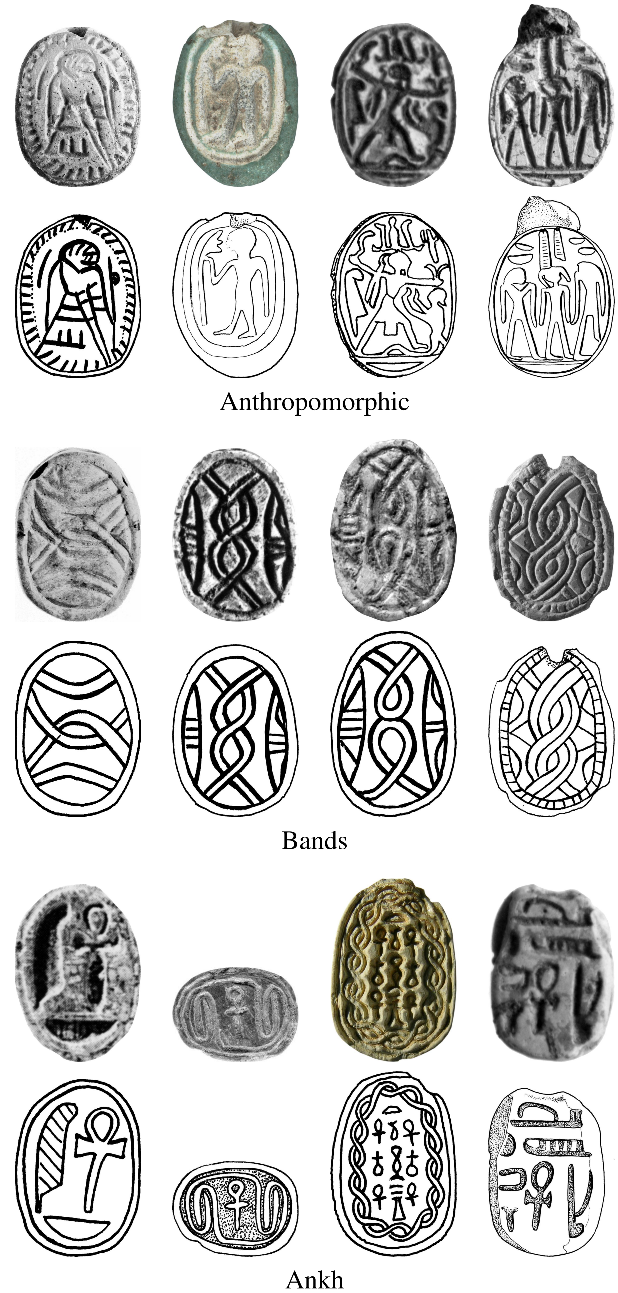


Fig. 17. **Samples of classes from CSSL dataset by shape: Bands, Anthropomorphic and Ankh.**

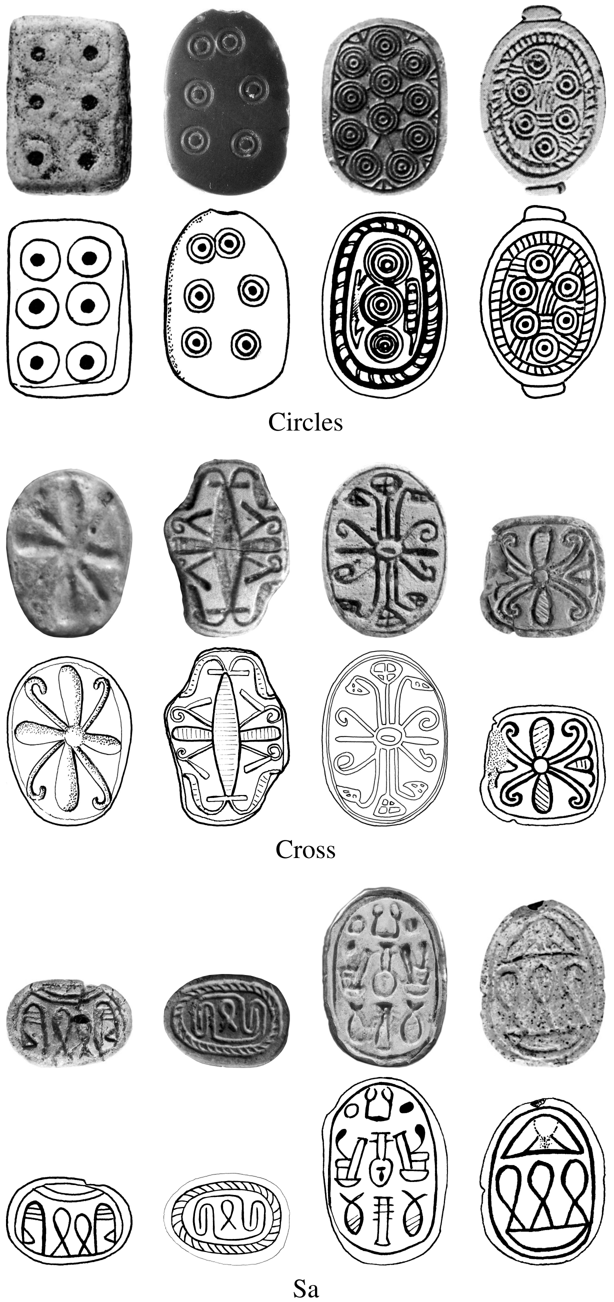


Fig. 18. **Samples of classes from CSSL dataset by shape: Circles, Cross and Sa.**

Fig. 19. **Sample of classes from CSSL dataset by sub-periods: Middle Bronze and Late Bronze.**

Iron: Early

Iron: Late

Fig. 20. **Sample of classes from CSSL dataset by sub-periods: Iron**