Teaching AI to Teach:
Leveraging Limited Human Salience Data Into Unlimited Saliency-Based Training

Colton Crum (presenting), Aidan Boyd, Kevin Bowyer, Adam Czajka
University of Notre Dame, {ccrum, aboyd3, kwb, aczajka}@nd.edu, https://cvrl.nd.edu

BMVC 2023

Human-guided Models Come at a Cost

Ongoing challenge
Human-guided training has proven to boost generalization capabilities of deep learning-based models. However, acquiring human annotations is costly.

Our solution
We utilize models first taught by human annotations ("Teacher models") to then train "Student models" through their saliency maps using the CYBORG loss.

CYBORG loss function

$$\mathcal{L} = \frac{1}{K}\sum_{k=1}^{K}\sum_{c=1}^{C}\mathbf{1}_{y_k \in \mathcal{C}_c}\Big[\underbrace{(1-\alpha)\|\mathbf{s}_k^{(\text{teacher})} - \mathbf{s}_k^{(\text{model})}\|^2}_{\text{teacher saliency loss component}} \underbrace{- \alpha \log p_{\text{model}}\big(y_k \in \mathcal{C}_c\big)}_{\text{classification loss component}}\Big]$$

$\|.\|$ = L_2 norm
y_k = class label for the k-th sample
$\mathbf{1}[.]$ = 1 when $y_k \in C_c$ (0 otherwise)
C = total number of classes
K = batch size
α = 0.5 (trade-off parameter)

Research Questions

- **(RQ1)** Which type of training produces better Teacher models: human-guided or purely data-driven?
- **(RQ2)** Can the top-performing Teacher model improve the performance of Student models across different CNN architectures?
- **(RQ3)** What are the potential performance benefits of the Teacher-Student training paradigm over the baselines?
- **(RQ4)** Can this training approach be applied to domains beyond synthetic face detection?

Once Taught By Humans, Models Can Teach Themselves

Table 2: Area Under the Curve (AUC, mean ± std) achieved by the baselines and the optimal student model in synthetic face detection task. Optimal AI Student configurations were achieved by training the model using the optimal AI Teacher's configuration (Xception + CAM) with $\alpha = 0.01$ (*i.e.* encouraging the model to focus on salience instead of class label).

Model	Baseline 1 (small set, entire human salience available)	Baseline 2 (larger set, no human salience)	Optimal AI Student (this paper: large set, optimal use of human salience)
DenseNet	0.633 ± 0.04	0.629 ± 0.039	**0.767 ± 0.020**
ResNet	0.612 ± 0.05	0.555 ± 0.061	**0.718 ± 0.012**
Xception	0.730 ± 0.02	0.586 ± 0.074	**0.743 ± 0.005**
Inception	0.679 ± 0.03	0.610 ± 0.035	**0.746 ± 0.019**

Teacher / Student Training Paradigm