Ultrasound Image Representation Learning by Modeling Sonographer Visual Attention

Richard Droste[1], Yifan Cai[1], Harshita Sharma[1], Pierre Chatelain[1], Lior Drukker[2], Aris T. Papageorghiou[2], J. Alison Noble[1]

[1]Department of Engineering Science, University of Oxford, UK [2]Nuffield Department of Women's & Reproductive Health, University of Oxford, UK

Research Question

Can transferable image representations be obtained automatically by learning to predict expert gaze?

Introduction

Motivation

- **Image representations** are commonly learned via class labels
  - Simplistic approximation of human image understanding
- Humans direct their **visual attention towards informative regions**
  - Location of visual attention can be recorded via gaze tracking
- The **visual patterns** that attract gaze can be **learned with CNNs**
  - "Visual attention model (VAM)" (visual saliency prediction)
- Can **visual attention models** transfer to **medical imaging tasks?**

Summary

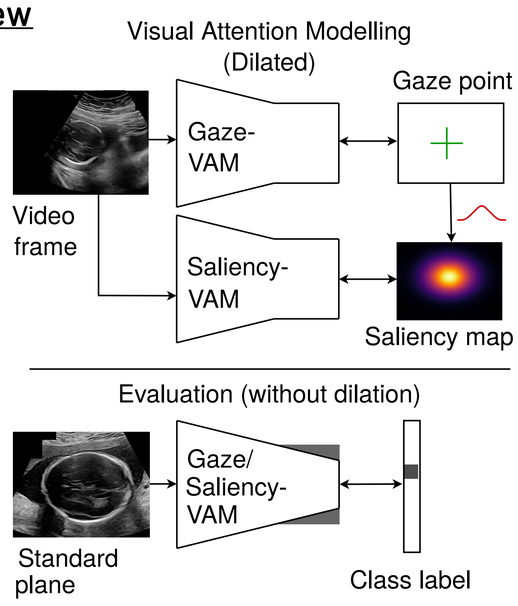
- **Image representations** are learned by training a CNN to predict **automatically acquired gaze** on routine ultrasound scans
- The learned representations are evaluated on the task of **detecting anatomical standard views** [1]

Method

Overview

1) **Visual Attention Modeling (VAM)**
   Train dilated CNN to predict operator gaze via:
   a) **Gaze point regression**
   b) **Visual saliency prediction**

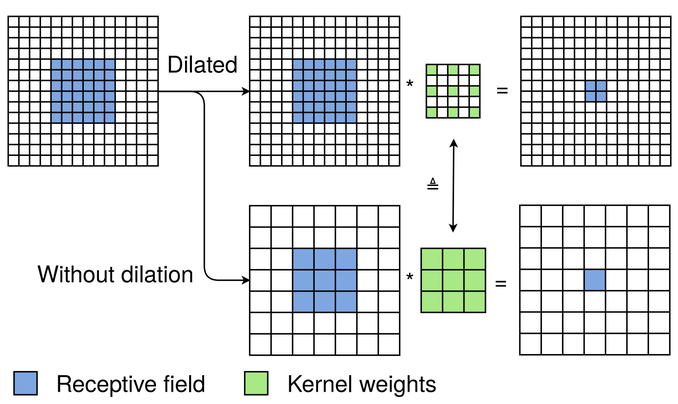
2) **Evaluation**
   Evaluate CNN without dilation on the task of **standard plane detection**



Dilated Convolutions

Visual attention modelling
1) Insert dilations
2) Remove downsampling
→ Preserve spatial information

Classification
1) Remove dilations
2) Insert downsampling
→ Preserve receptive field



Visual Saliency Prediction

- Predicts **gaze point probability distribution** $\hat{\mathbf{S}}$ from activations $\mathbf{A}$

$$\hat{S}_{i,j} = e^{A_{i,j}} / \sum_{i,j} e^{A_{i,j}}$$

- Target $\mathbf{S}^*$ is generated as mixture of Gaussians around gaze points
- Kullback-Leiber divergence loss

$$\mathcal{L}_s(\mathbf{S}^*, \hat{\mathbf{S}}) = D_{\mathrm{KL}}(\mathbf{S}^* \| \hat{\mathbf{S}})$$

Gaze Point Regression

- Model regresses geometric median of gaze points via **soft-argmax** [2]

$$\hat{\mathbf{p}} = \sum_{i,j} \hat{S}_{i,j} \left[\frac{j-0.5}{W_D} W,\ \frac{i-0.5}{H_D} H\right]^\top$$

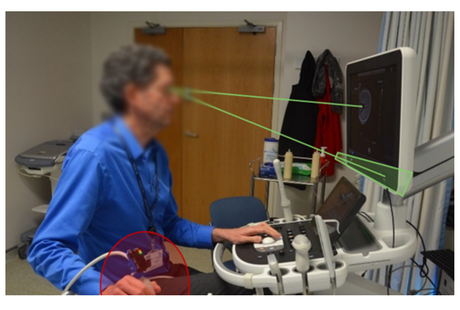
- L2 loss

$$\mathcal{L}_g(\mathbf{p}^*, \hat{\mathbf{p}}) = \|\mathbf{p}^* - \hat{\mathbf{p}}\|_2$$

Data

The PULSE project

Perception Ultrasound by Learning Sonographic Experience:
- Full-length videos of fetal ultrasound scans
- Simultaneous recording of operator gaze tracking and probe motion data
- To better understand and facilitate ultrasound



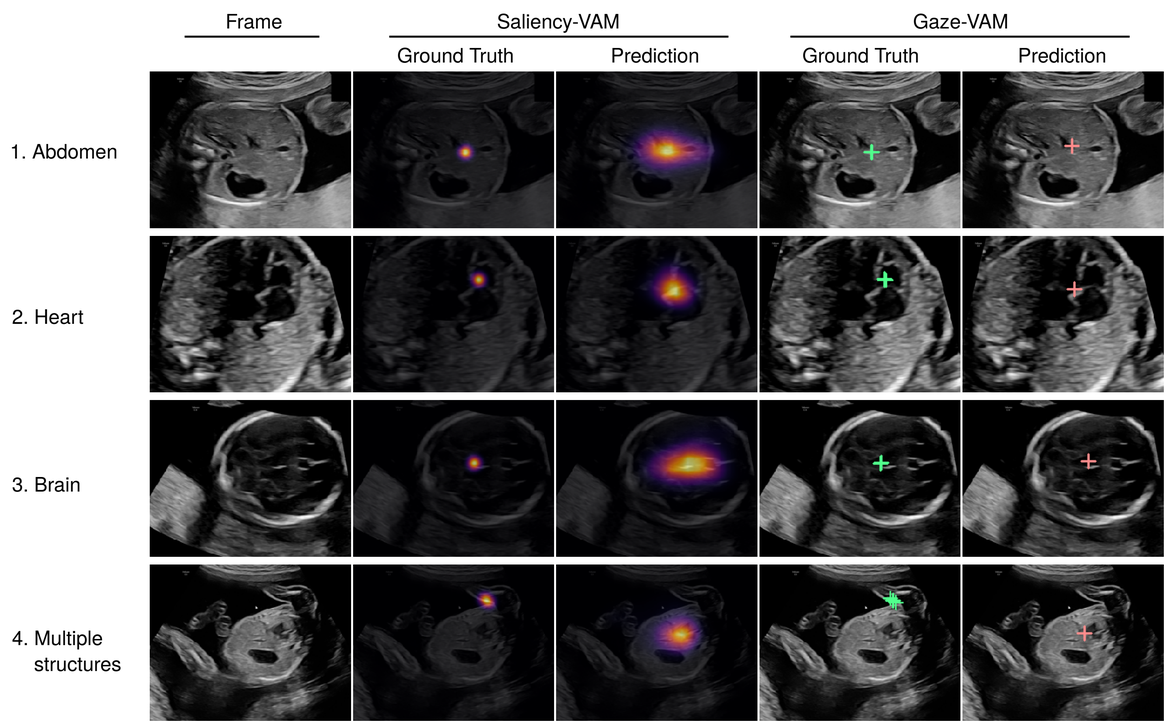
Visual Attention Modeling

Quantitative Evaluation

	Saliency-VAM					Gaze-VAM
	KLD	NSS	AUC [%]	CC [%]	SIM [%]	ℓ_2-norm
Static	3.41 ±0.02	1.39 ±0.05	85.9 ±0.3	14.9 ±0.4	8.5 ±0.1	54.4 ±0.6
Learned	**2.43** ±0.03	**4.03** ±0.05	**96.7** ±0.2	**31.6** ±0.3	**18.5** ±0.2	**27.4** ±0.4

- Predictions more accurate than typical benchmark scores and related work [3]
- Lack of additional baseline methods since typical saliency predictors cannot handle classification

Qualitative Evaluation



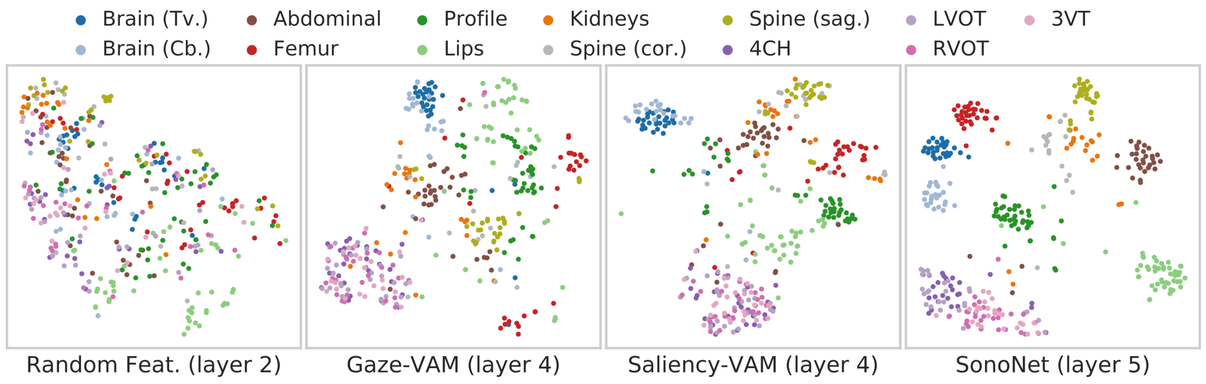
Standard Plane Detection

Transfer Learning

	Rand. Init.	Gaze-FT	Saliency-FT	Δ(Saliency, Rand. Init.)	SonoNet-FT (Lit. value [1])
Precision	70.4 ± 2.3	67.2 ± 3.4	**79.5** ± 1.7	9.1 ± 2.1	**82.3** ± 1.3 (81)
Recall	64.9 ± 1.6	57.3 ± 4.5	**75.1** ± 3.4	10.2 ± 1.9	**87.3** ± 1.1 (86)
F1-score	67.0 ± 1.3	60.7 ± 3.9	**76.6** ± 2.6	9.6 ± 2.1	**84.5** ± 0.9 (83)

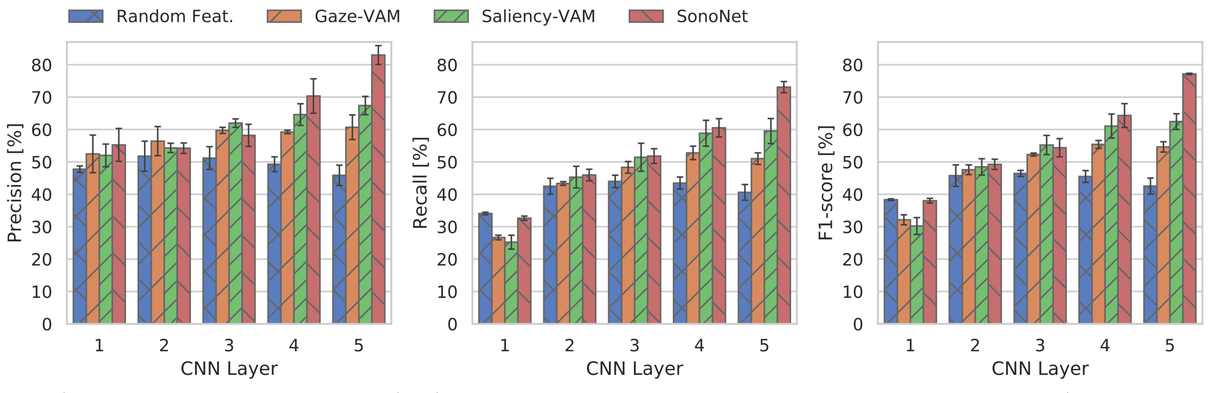
- **Significant improvement over training from random initialization**
- **Approaching fully supervised reference despite 20x less labeled data**

T-SNE Visualization of Feature Space



- Most standard planes are well-separated in feature space
- Overlap remains among views of the fetal heart and head

Regression on Fixed Representations



- High-level features predictive for fetal anomaly standard plane detection
- Predictiveness decreases in last layer, indicating task-specificity

References
[1] Baumgartner et al.: SonoNet: Real-Time Detection and Localisation of Fetal Standard Scan Planes in Freehand Ultrasound. IEEE Trans. Med. Imag. 36(11), 2204–2215 (2017).
[2] Levine, S., Finn, C., Darrell, T., Abbeel, P.: End-to-End Training of Deep Visuomotor Policies. J. Mach. Learn. Res. 17(1), 1334–1373 (2015).
[3] Cai et al.: Multi-task SonoEyeNet: Detection of Fetal Standardized Planes Assisted by Generated Sonographer Attention Maps. In: MICCAI (2018).