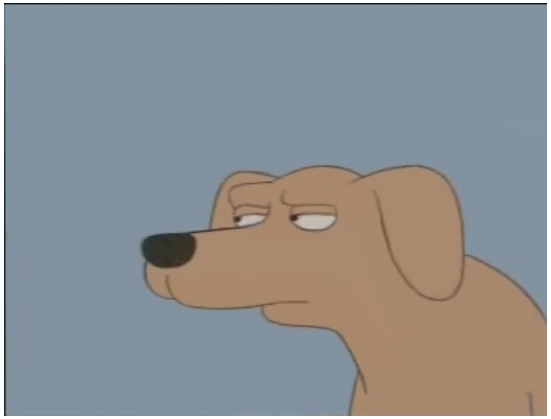
Where are they looking?



But why?

trefoils

Goal

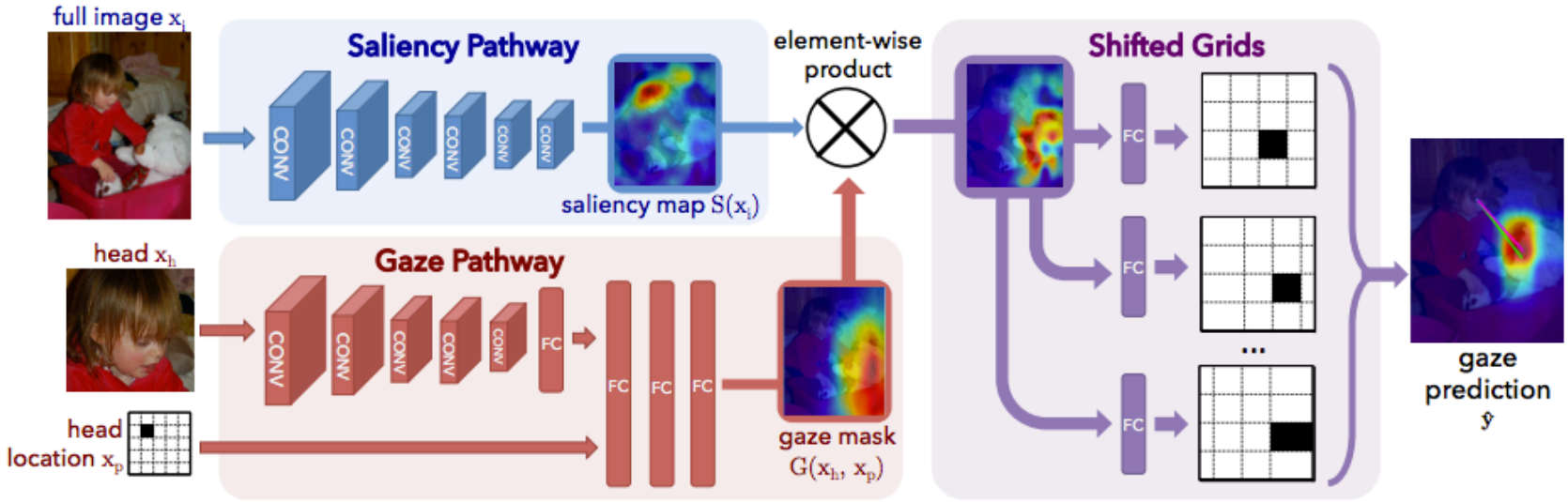
- Perform gaze-following in natural settings.
- Only a single view available.
- Without restrictive assumptions.

Related work

- Saliency:
  - Mainly fixations of an observer free-viewing pictures.
- Gaze:
  - Only estimate gaze direction.
  - Rely on face detection (so only work for people facing the camera).

GazeFollow: the dataset

ENOUGH SETUP
SHOW ME HOW THEY DID IT!!!
imgflip.com

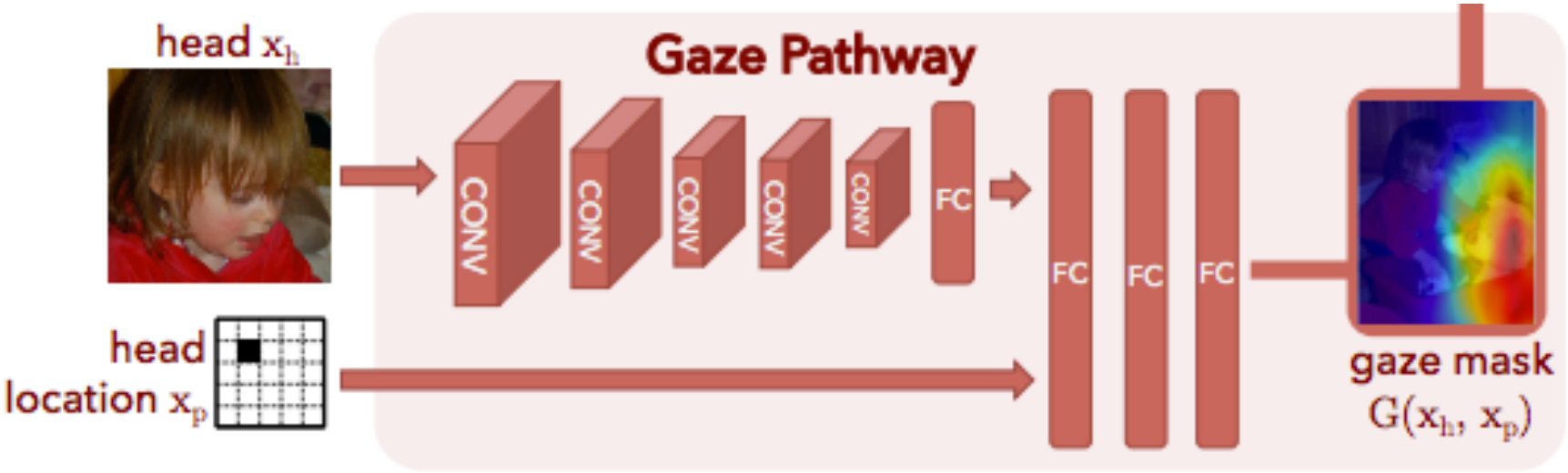


$$\hat{y} = F\left(G(x_h, x_p) \otimes S(x_i)\right)$$

Saliency map

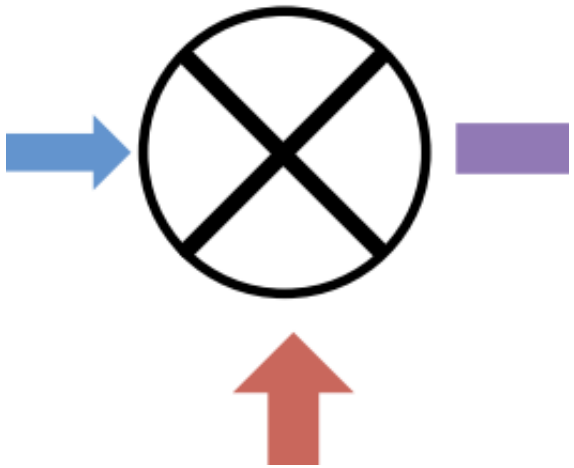
$$S(x_i)$$

Gaze mask



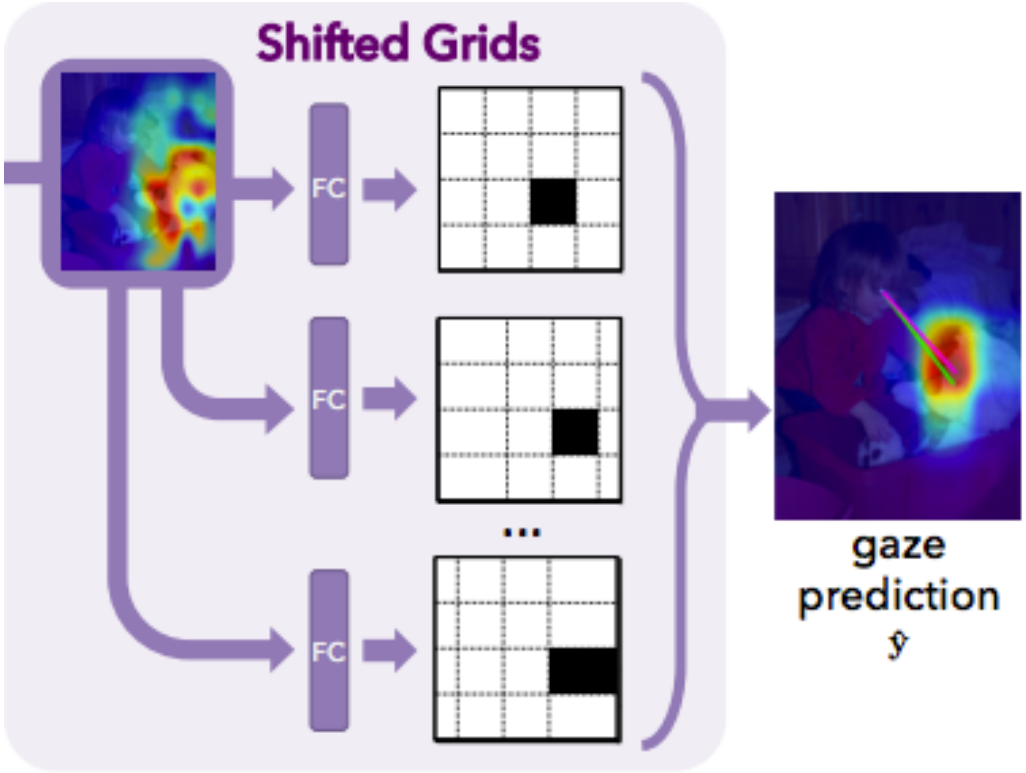
$$G(x_h, x_p)$$

element-wise
product



$$G(x_h, x_p) \otimes S(x_i)$$

Shifted grids



$$F\left(G(x_h, x_p) \otimes S(x_i)\right)$$

Quantitative results (pears to apples)

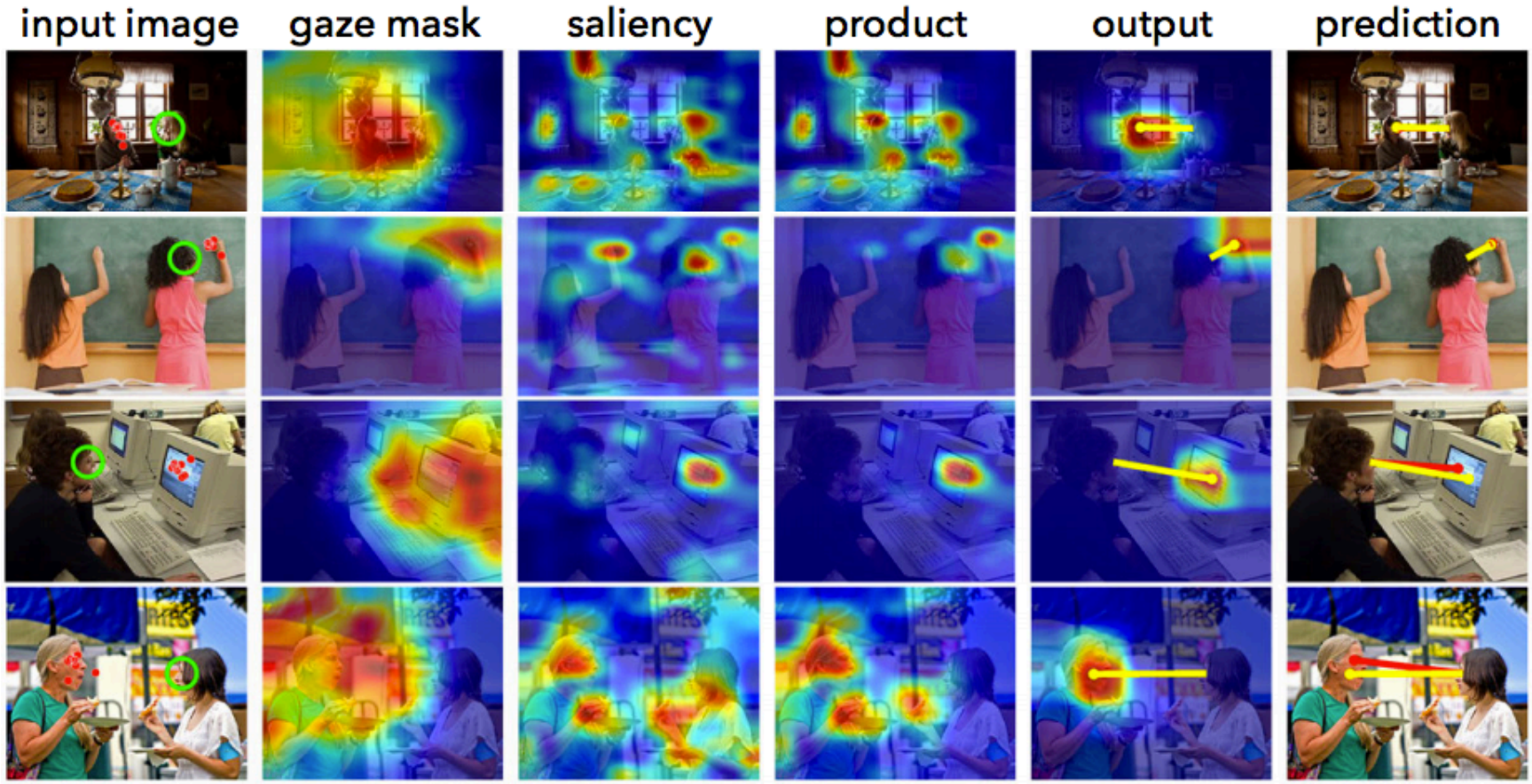
Model	AUC	Dist.	Min Dist.	Ang.
Our	**0.878**	**0.190**	**0.113**	**24°**
SVM+shift grid	0.788	0.268	0.186	40°
SVM+one grid	0.758	0.276	0.193	43°
Judd [11]	0.711	0.337	0.250	54°
Fixed bias	0.674	0.306	0.219	48°
Center	0.633	0.313	0.230	49°
Random	0.504	0.484	0.391	69°
One human	0.924	0.096	0.040	11°

Ablation study (removing stuff)

Model	AUC	Dist.	Min Dist.	Ang.
No image	0.821	0.221	0.142	27°
No position	0.837	0.238	0.158	32°
No head	0.822	0.264	0.179	41°
No eltwise	0.876	0.193	0.117	25°
5×5 grid	0.839	0.245	0.164	36°
10×10 grid	0.873	0.218	0.138	30°
L2 loss	0.768	0.245	0.169	34°
Our full	0.878	0.190	0.113	24°

Qualitative Results
(pew pew pew)

Visualization of internal representations



Critique

Negatives:

- Not using “just one image as input” (but they do in the demo).
- In the quantitative analysis, they compare with methods that solve different problems.

Critique

Positives:

- Awesome new dataset.
- Nice first attempt (and baseline for future work).
- Lots of room to improve.
- Great as a future project.

Thanks!