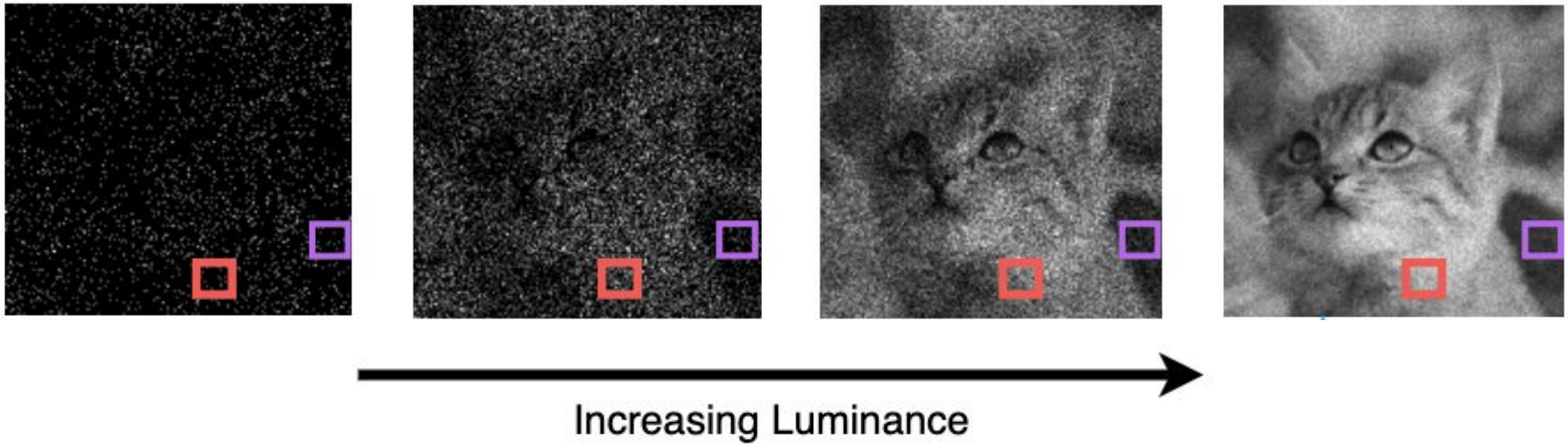
Seeing Into Darkness: Scotopic Visual Recognition

By Bo Chen and Pietro Perona

Yash Sanghvi
ML Reading Group
October 30th, 2019

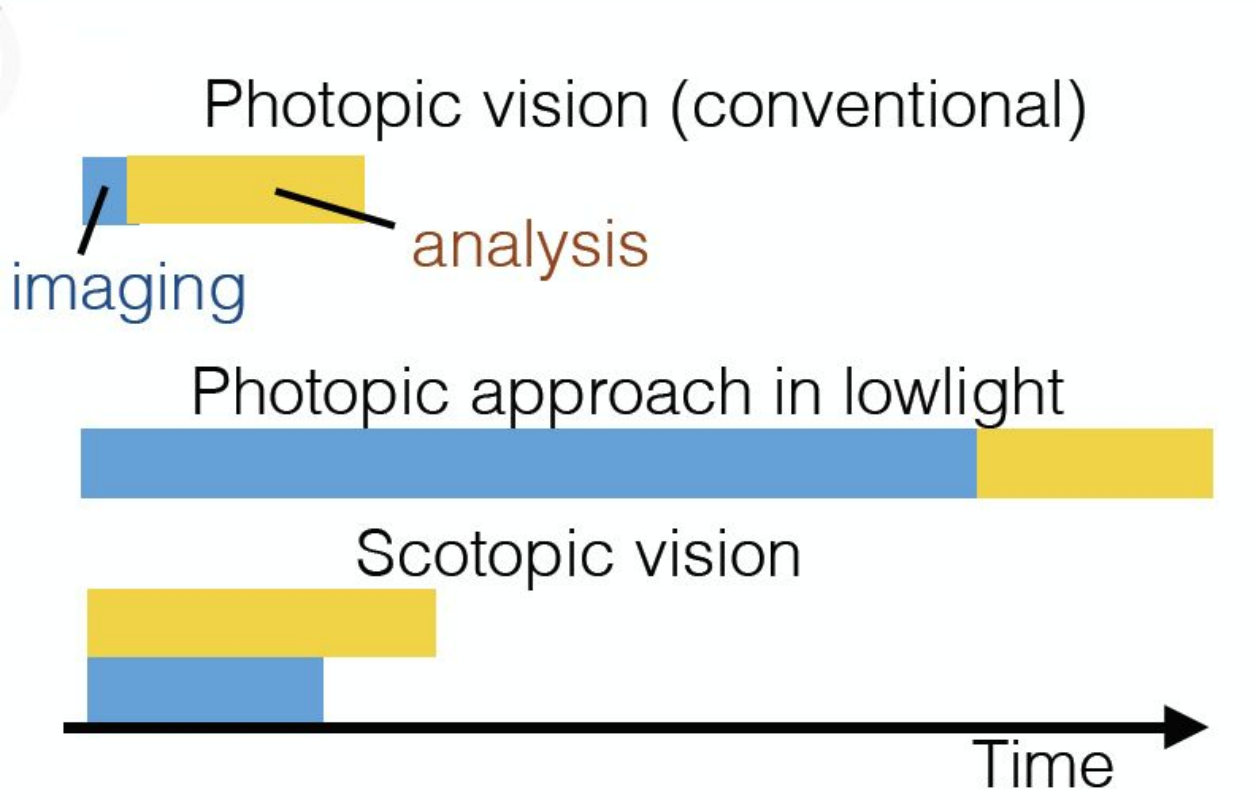
Scotopic: *Vision in the Dark*



PPP: Photons Per Pixel

PPP $\propto$ Exposure Time

What's Different from Photopic Vision?



Some Mathematical Formulation

$X_{1:t}$: Stream of Photons incident on sensors

Photons incident on pixel ‘i’ during time bin ‘t’

$$P(X_{t,i} = k) = Poisson(k|\lambda_i t)$$

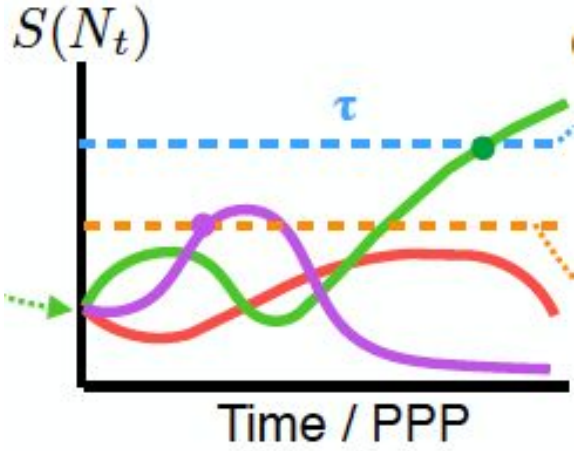
Classification with Scotopic Vision

Sequential Probability Ratio Test

$$\text{Compute } c^* = \arg \max_{c=1,2,\ldots,} S_C(X_{1:t})$$

$$\text{if } S_{c^*} > \tau : \text{ report } \hat{C} = c^*$$

$$\text{otherwise increase exposure time } t$$



WaldNet

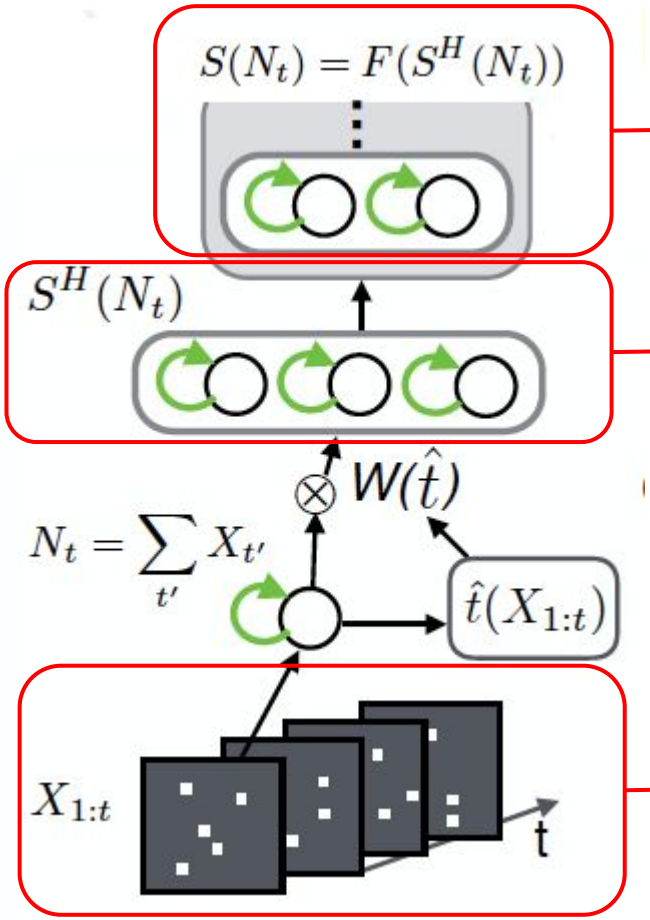
- Standard ConvNets not applicable
- Instead work with High Exposure Time ‘T’
- Working with partial observations $N_t = \sum_{t'=1}^{t} X_{t'}$
- Uncounted Photons $\Delta N = \sum_{t'=t+1}^{T} X_{t'}$

Gamma Prior on photon emission rate: $Gam(\mu_i t_0, t_0)$

First “Convolutional“ Layer

$$S^H(\boldsymbol{N}_t) = \sum_{\Delta \boldsymbol{N}} (\boldsymbol{W}(\boldsymbol{N}_t + \Delta \boldsymbol{N}) + \boldsymbol{b}^H) P(\Delta \boldsymbol{N} | \boldsymbol{N}_t) \approx \alpha(t) \boldsymbol{W} \boldsymbol{N}_t + \boldsymbol{\beta}(t) \qquad (4)$$

$$\alpha(t) \triangleq \frac{T+t_0}{t+t_0} \qquad \beta_j(t) \triangleq \frac{t_0(T-t)}{t+t_0} \sum_i W_{ij}\mu_i + b_j$$



Calculation of Log Posterior Probability

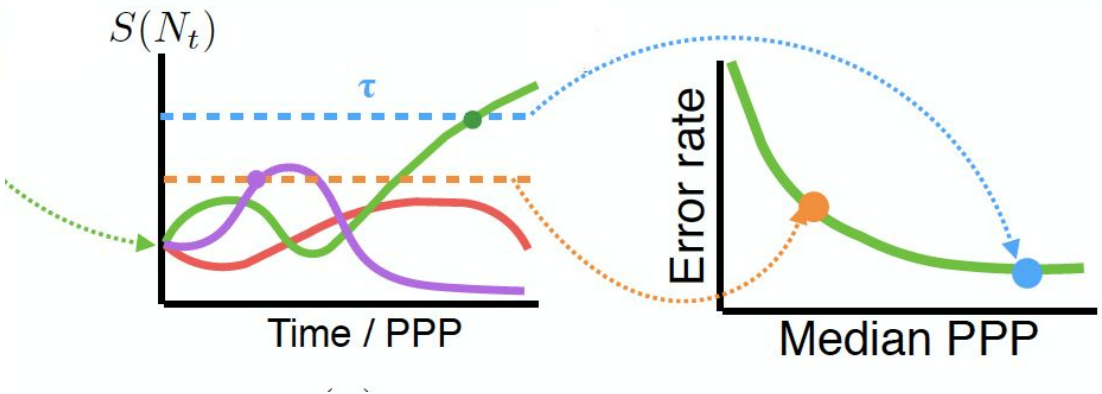
Time Invariant Features

Stream of Photons

Learning

$$R \triangleq \eta \mathbb{E}[\mathrm{PPP}] + \mathbb{E}[C \neq \hat{C}]$$

Bayes Risk



Two Stage Learning

- (I) Learning ConvNet/Waldnet
- (II) Threshold Learning

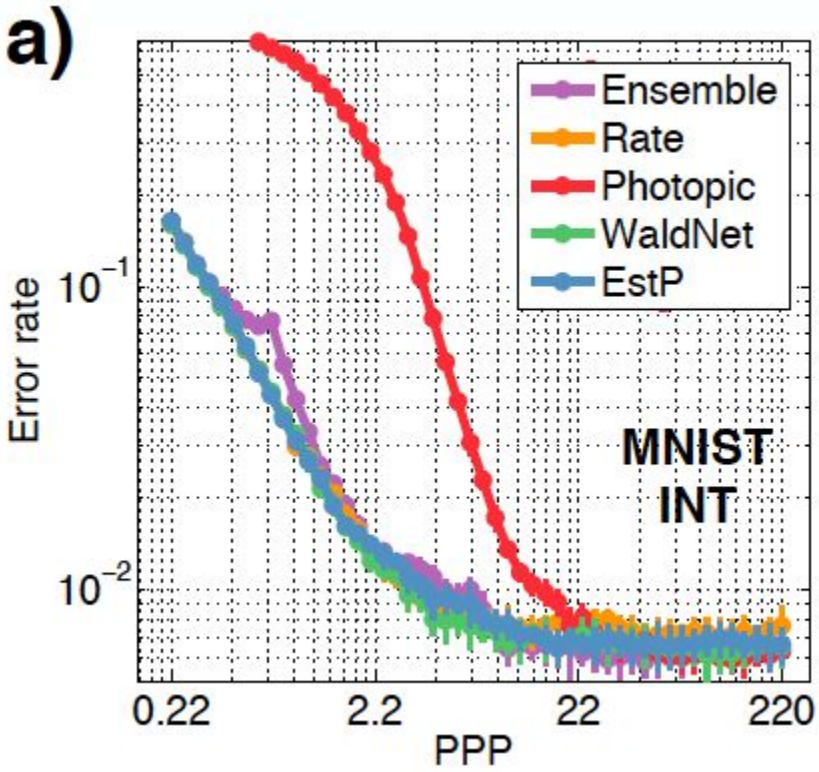
Threshold Learning

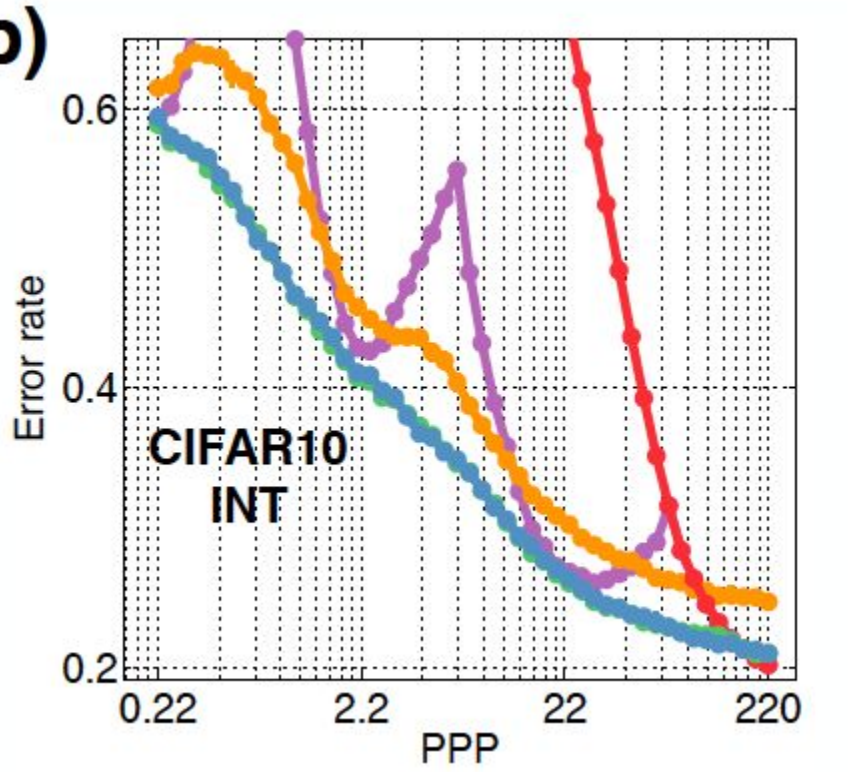
Bayes Risk at time t

$$R_t^{(n)} = \eta\Delta + q_t^{(n)} e_t^{(n)} + (1 - q_t^{(n)}) R_{t+1}^{(n)}$$

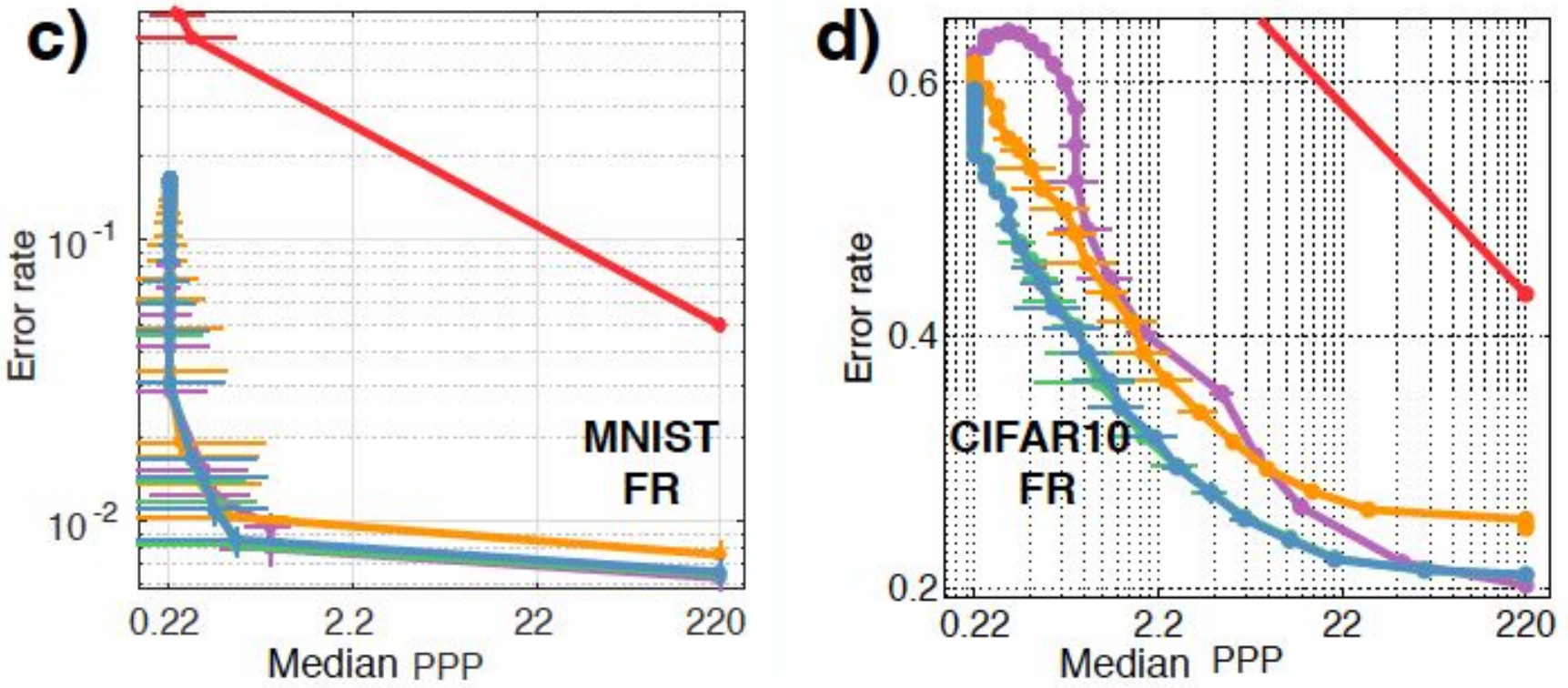
$$\text{Bayes Risk: } R = E_{(n)}[R_0^{(n)}]$$

Results - Interrogation Regime

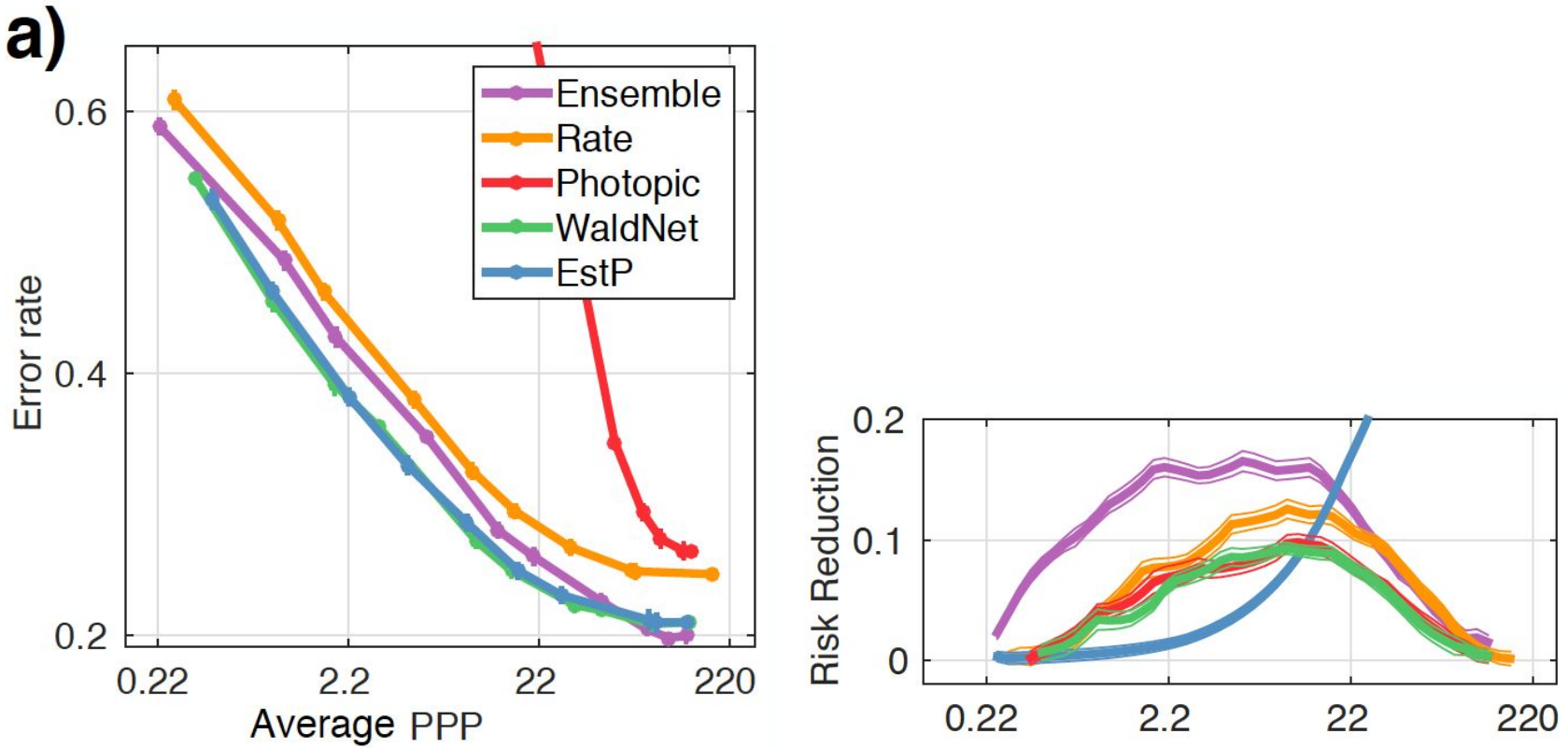




Results - Free Response Regime



Risk Reduction with Threshold Learning



Conclusion

- *WaldNet* for Scotopic Vision, Designed to process images as they hit the photoreceptors

- Provide a transform to convert stream of photons to “time invariant features”

$$S^H(\boldsymbol{N}_t) = \sum_{\Delta \boldsymbol{N}} (\boldsymbol{W}(\boldsymbol{N}_t + \Delta \boldsymbol{N}) + \boldsymbol{b}^H) P(\Delta \boldsymbol{N} | \boldsymbol{N}_t) \approx \alpha(t) \boldsymbol{W} \boldsymbol{N}_t + \beta(t) \qquad (4)$$

- Allows tradeoff between computational time vs. accuracy.

*Did not talk about Sensor noise and Spiking Implementation in this presentation