Dr. B Panos

Flare prediction with explainable artificial intelligence

SCOSTEP

16 May 2023

Prof. L Kleint

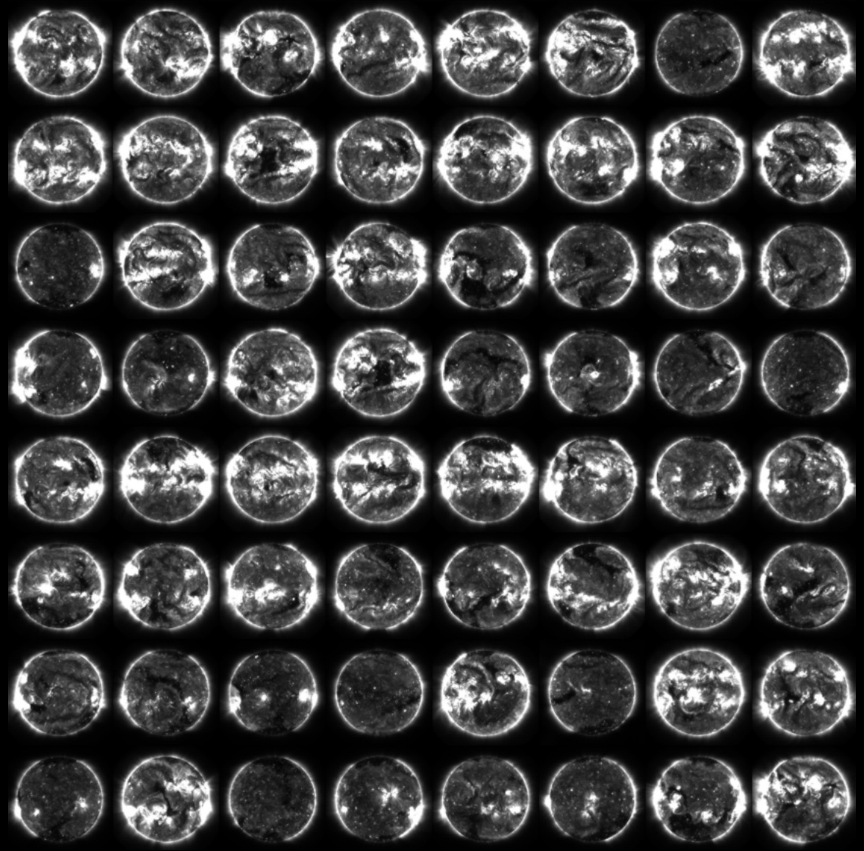
J. Zbinden

UNIVERSITÄT BERN

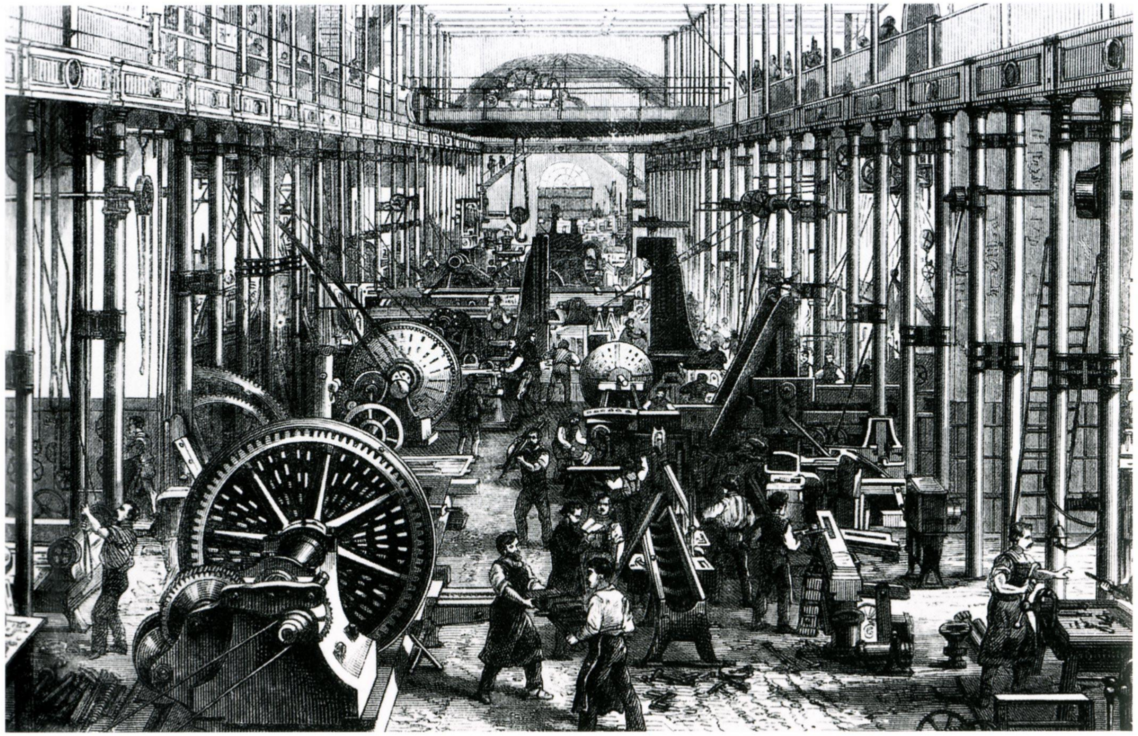
UNIVERSITY OF BERN

MNIST

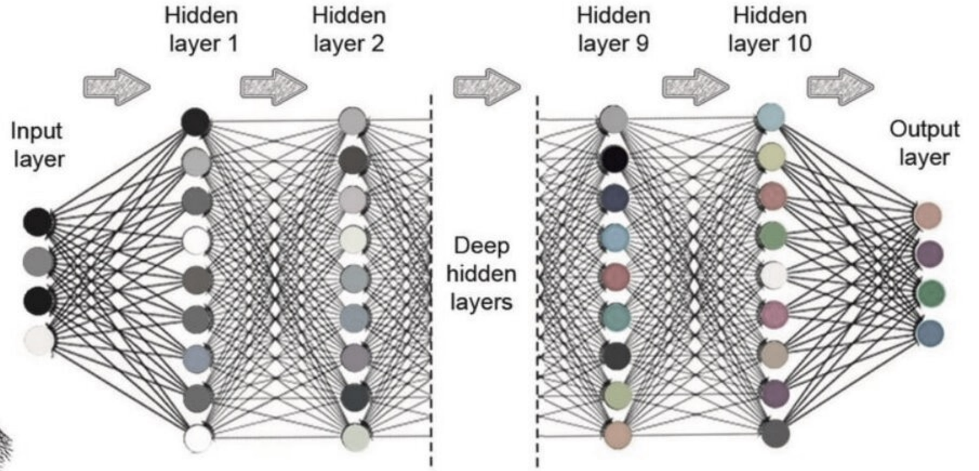
NASA/SDO



These techniques are data agnostic, so we can use them to solve our open questions

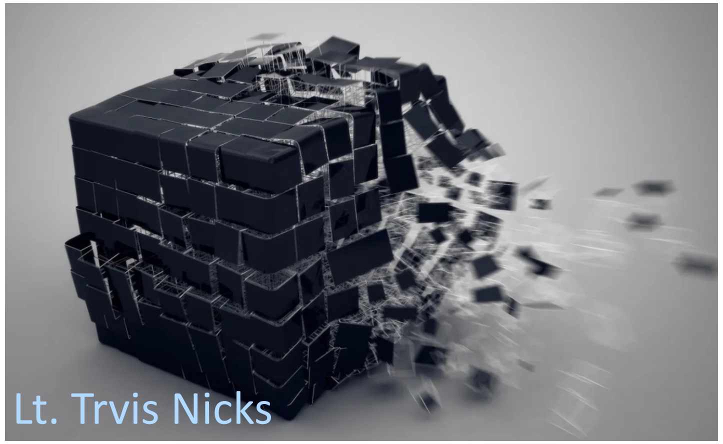


ML was nurtured by industry, where outcome supersedes understanding



P. G. Breen et.al MNRAS 2020, illustration by Mahala Le May

The deep learning revolution further reduced the role of understanding

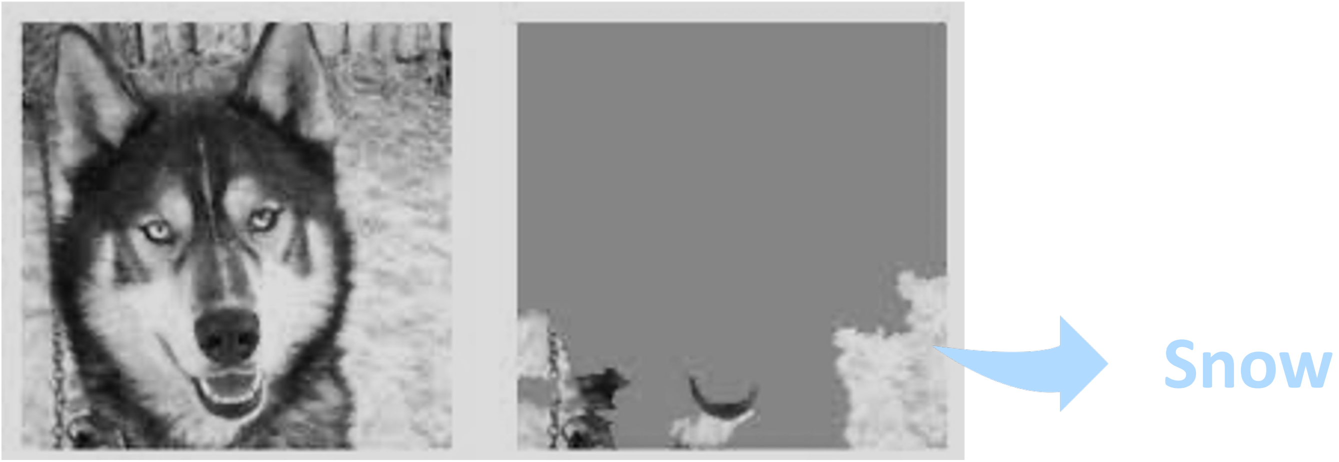


Lt. Trvis Nicks

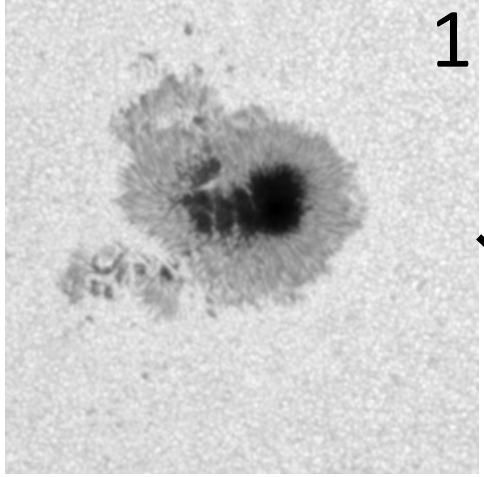
Neural networks effectively function as efficient black boxes

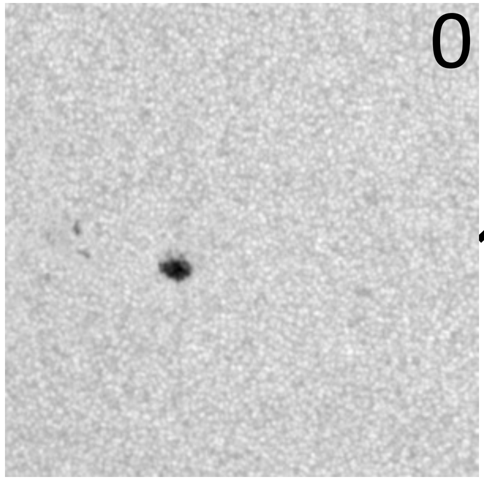
Ribeiro et. al, *Why Should I Trust You?* (2016)

Ribeiro et. al, *Why Should I Trust You?* (2016)

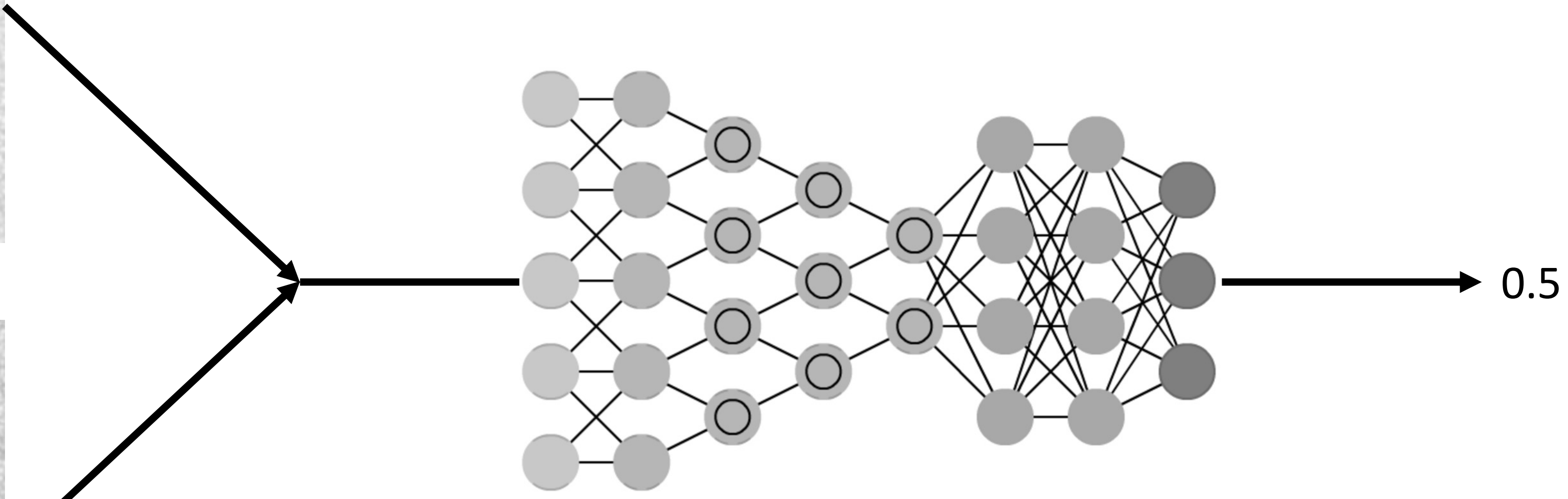


Sunspot





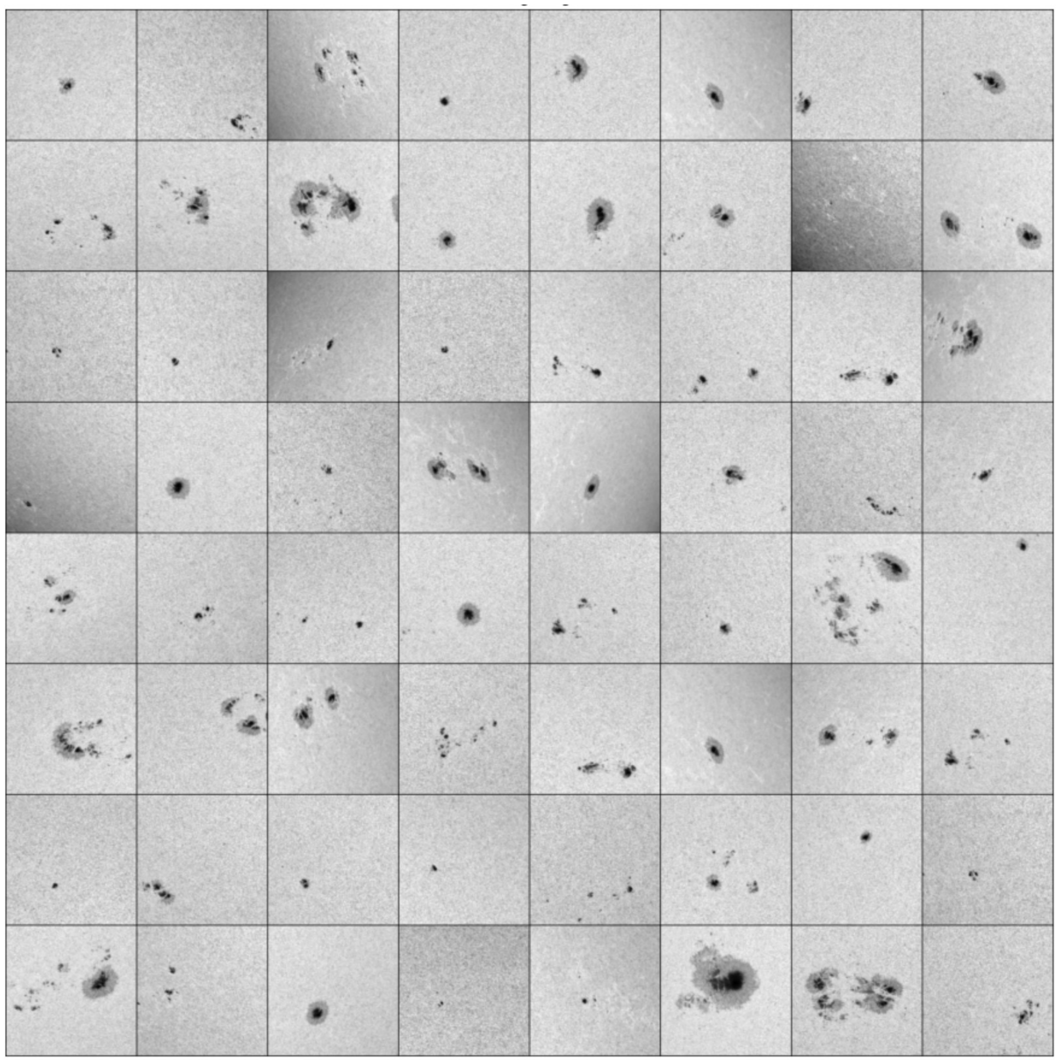
Pores

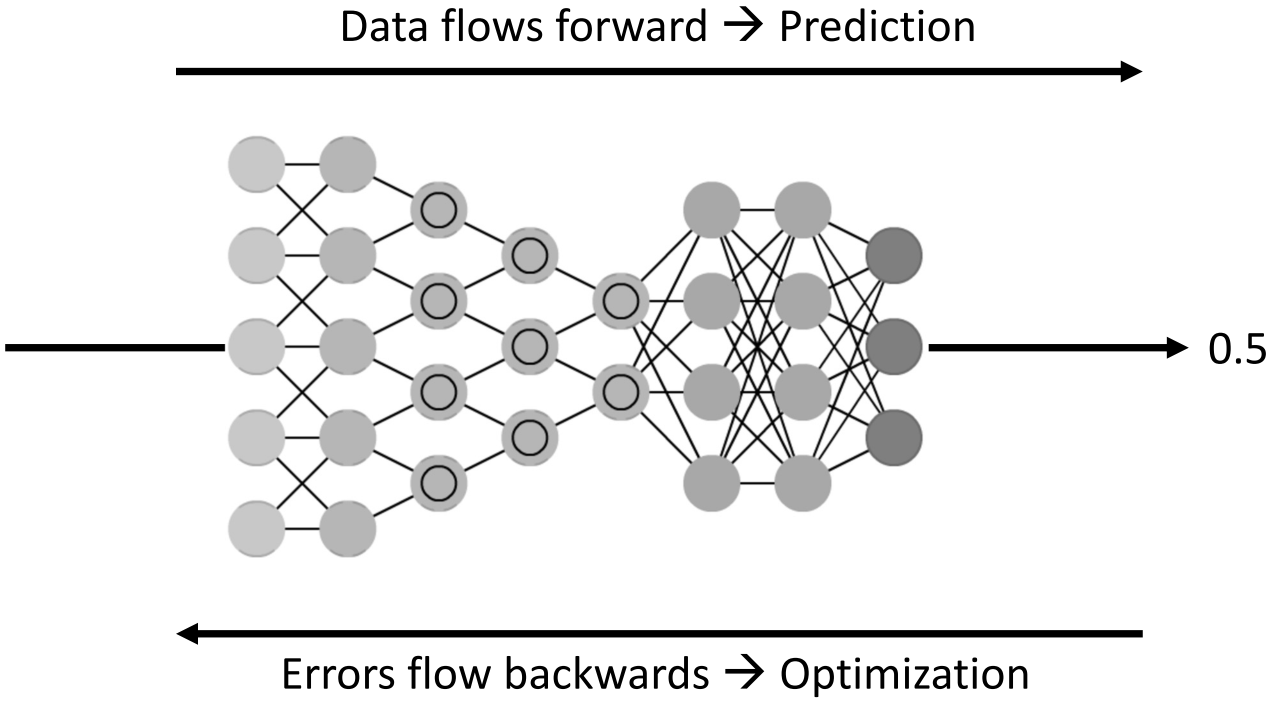


$$\mathcal{F}_{\Theta}(x)$$

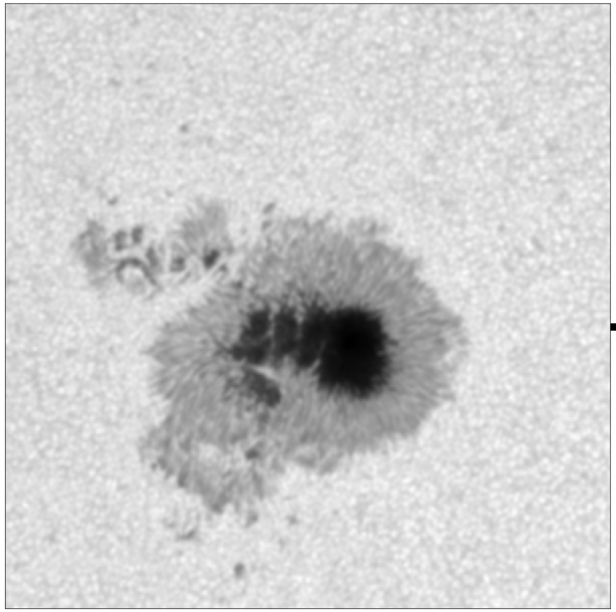
1. Function is a graph Like structure
2. Variational (start with bad guess)
3. Optimize via gradient descent using backpropagation

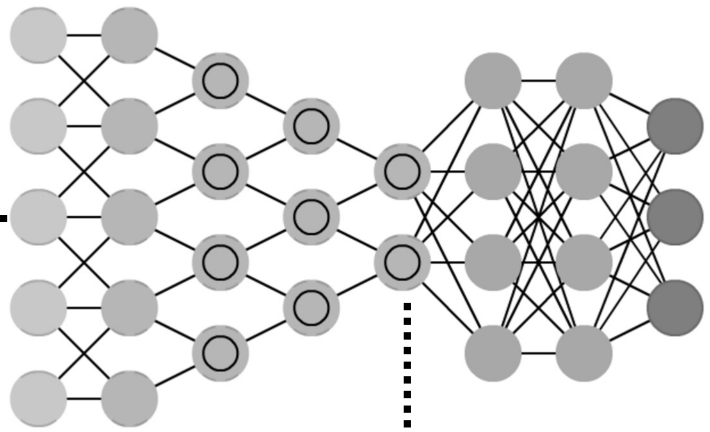
NASA/SDO




Data flows forward → Prediction
0.5
Errors flow backwards → Optimization

NASA/SDO

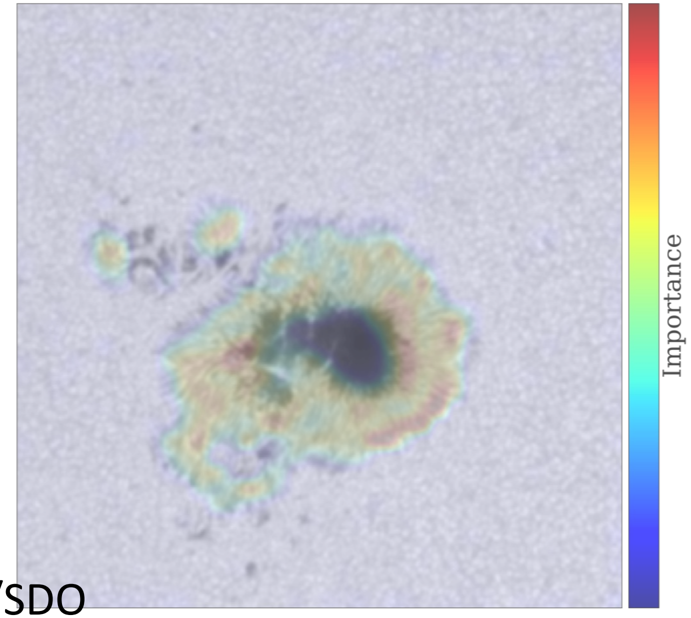




0.9

Explainable AI

Grad-CAM (Selvaraju et al. 2017)



Grad-CAM allows us to see what distinguishes a sunspot from a pore

Penumbra!

NASA/SDO

Reality basis

$$\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_k}A_i - \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_i}A_k + \frac{1}{c}\frac{\partial}{\partial x_k}\frac{\partial\phi}{\partial t} + \frac{1}{c^2}\frac{\partial^2 A_k}{\partial t^2} = \frac{4\pi}{c}J_k$$

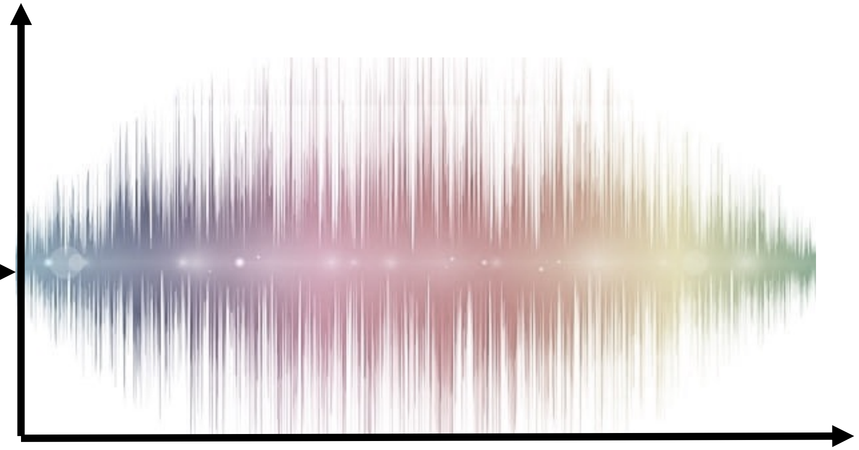
$$\frac{\partial}{\partial x_k}\vec{\nabla}\cdot\vec{A} - \nabla^2 A_k + \frac{1}{c}\frac{\partial}{\partial x_k}\frac{\partial\phi}{\partial t} + \frac{1}{c^2}\frac{\partial^2 A_k}{\partial t^2} = \frac{4\pi}{c}J_k$$

$$-\nabla^2 A_k + \frac{1}{c^2}\frac{\partial^2 A_k}{\partial t^2} + \frac{\partial}{\partial x_k}\left(\vec{\nabla}\cdot\vec{A} + \frac{1}{c}\frac{\partial\phi}{\partial t}\right) = \frac{4\pi}{c}J_k$$

$$-\nabla^2\vec{A} + \frac{1}{c^2}\frac{\partial^2\vec{A}}{\partial t^2} + \vec{\nabla}\left(\vec{\nabla}\cdot\vec{A} + \frac{1}{c}\frac{\partial\phi}{\partial t}\right) = \frac{4\pi}{c}\vec{J}$$

Human Brain

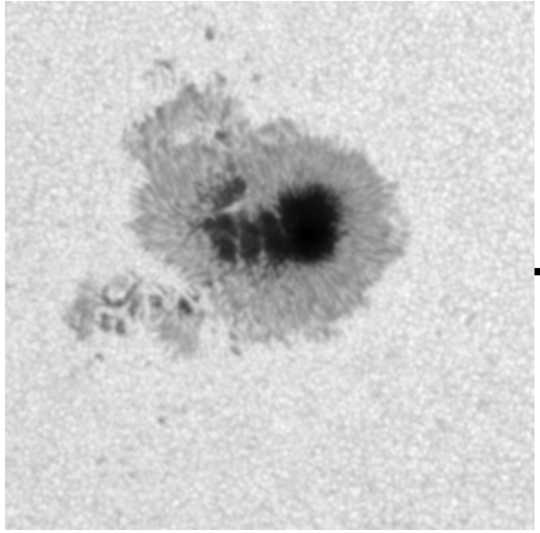
Human basis



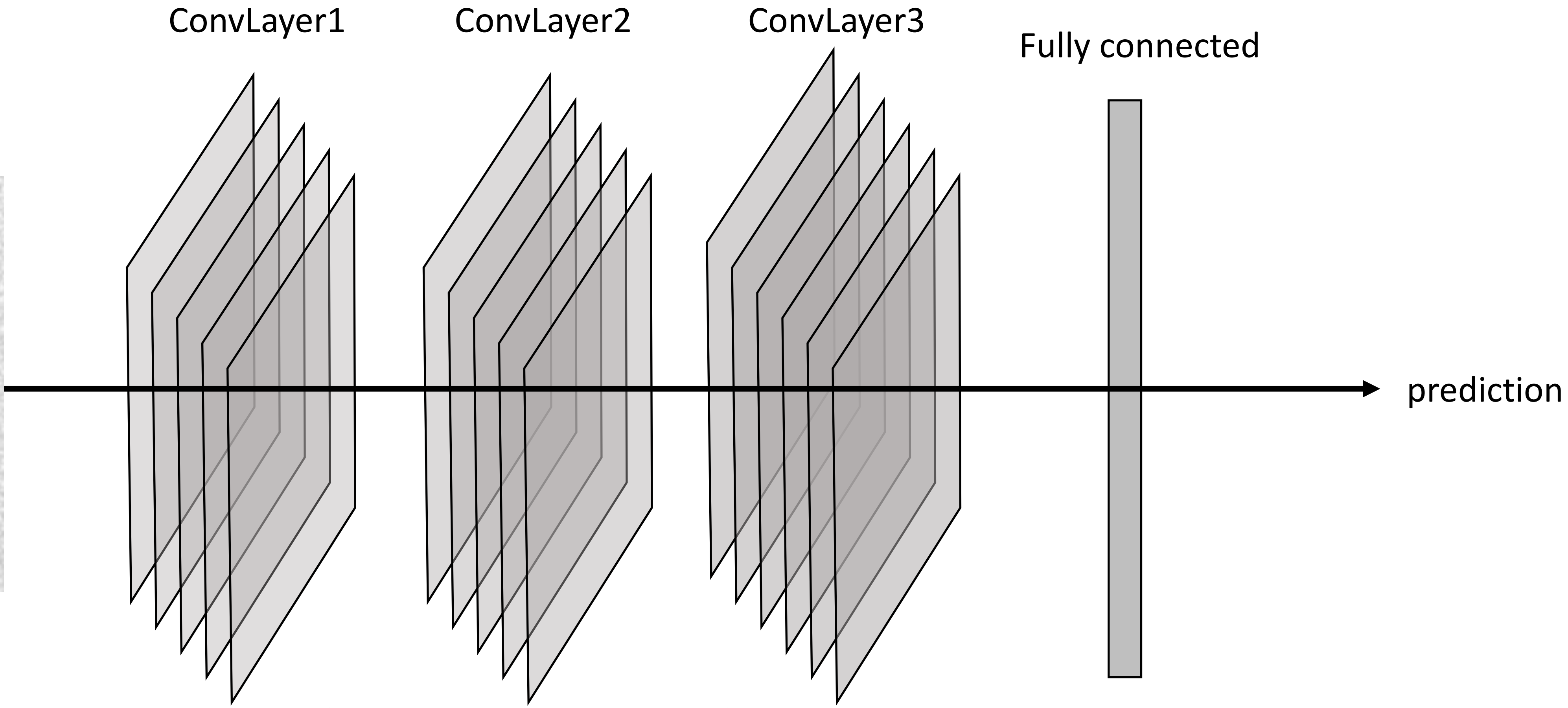
Maxwell's equations

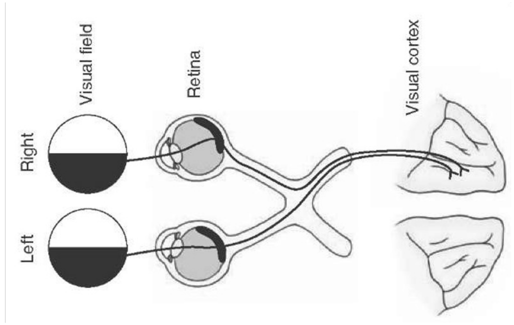
Abstraction

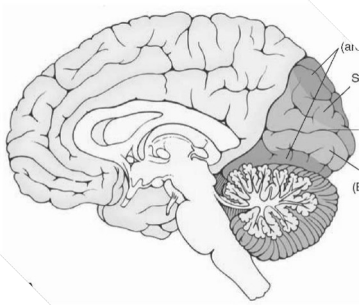
More useful basis of colors

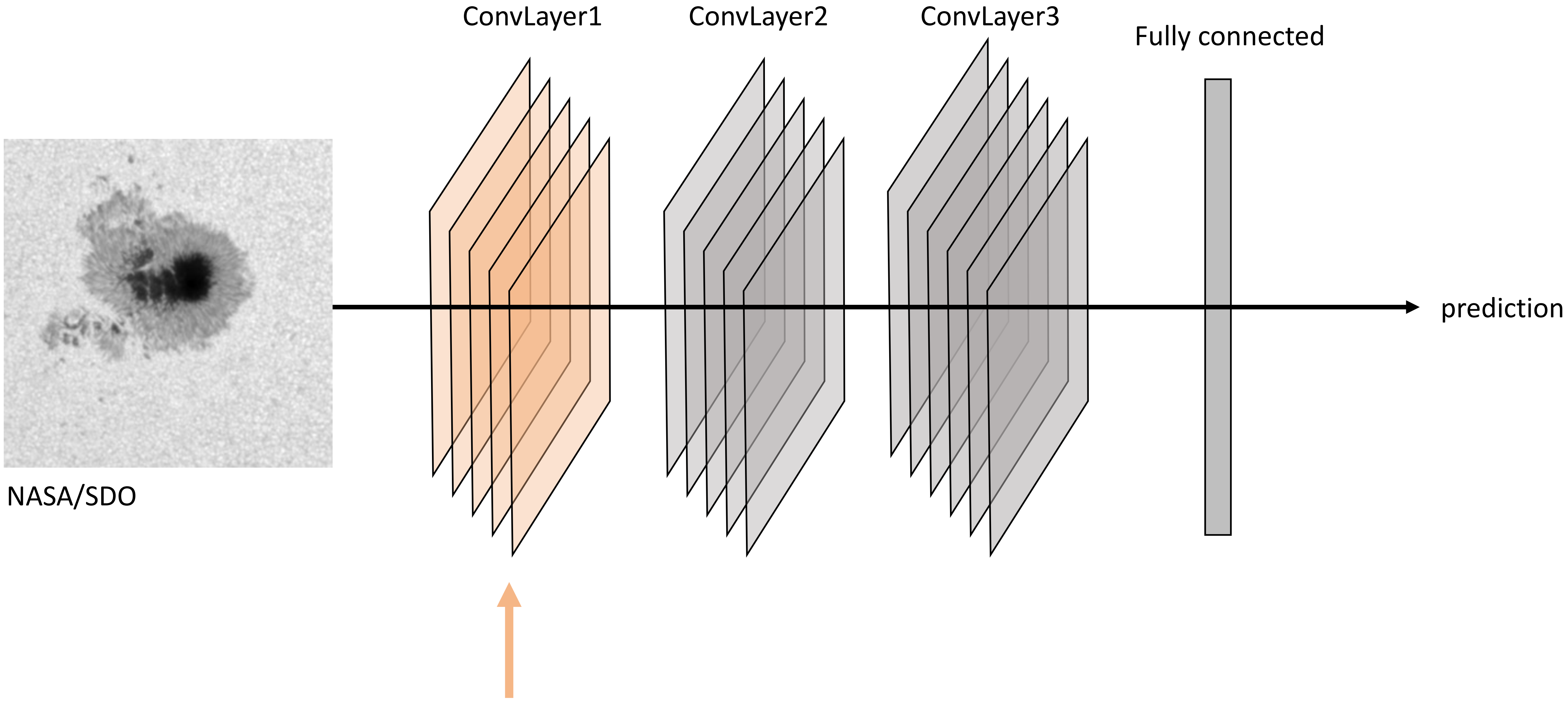


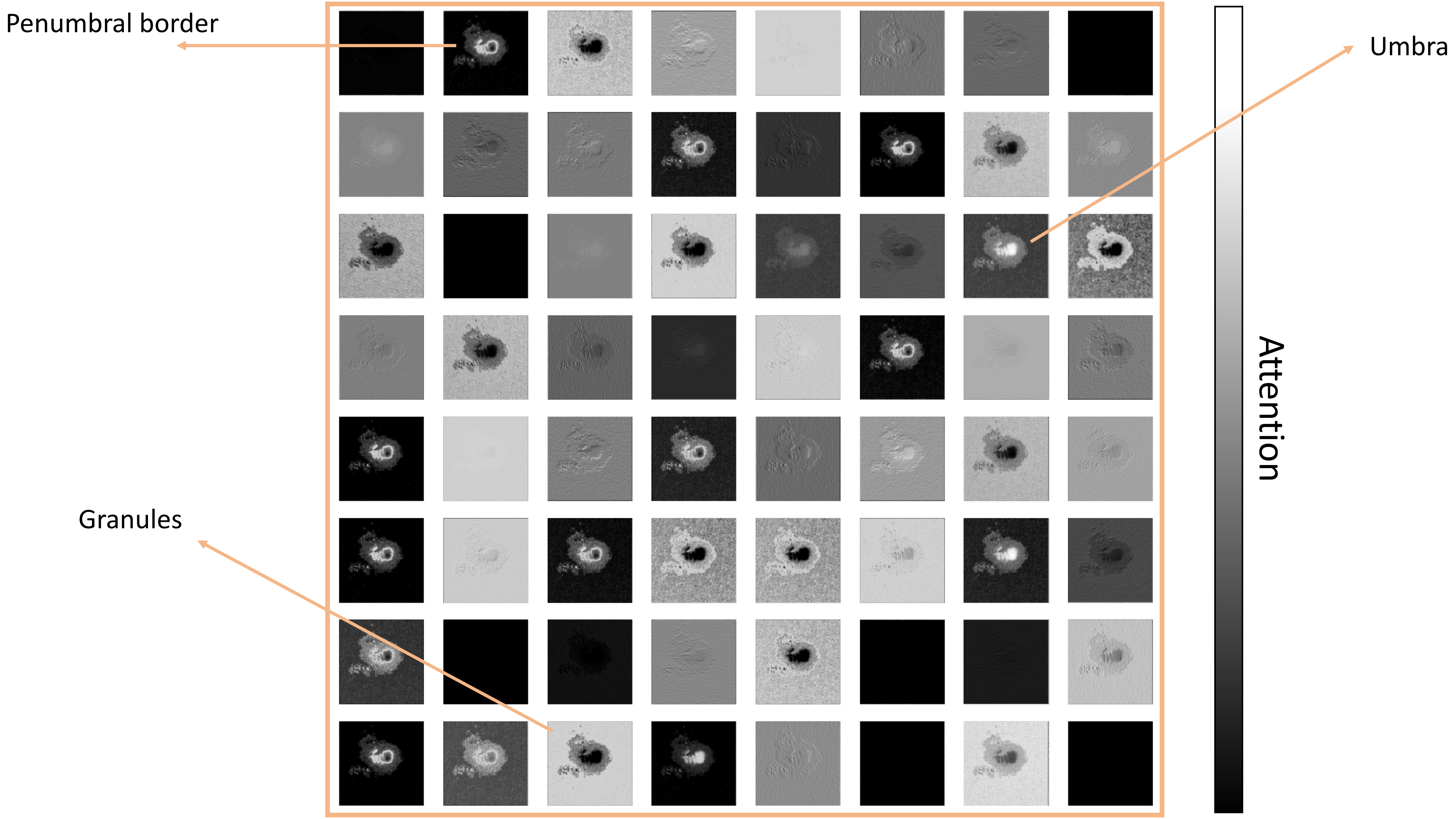
NASA/SDO

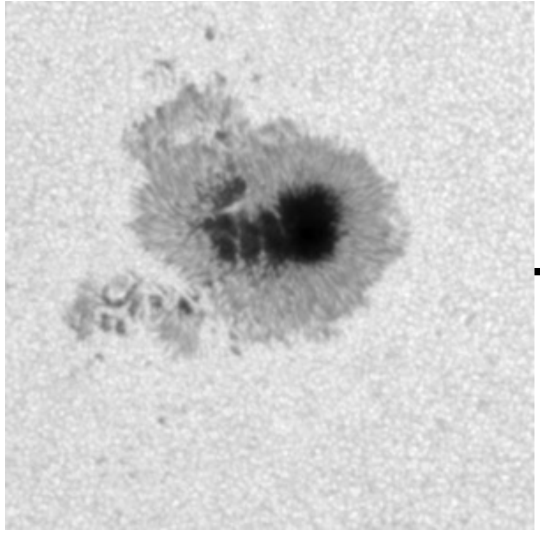

ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction


Visual field
Retina
Visual cortex
Right
Left

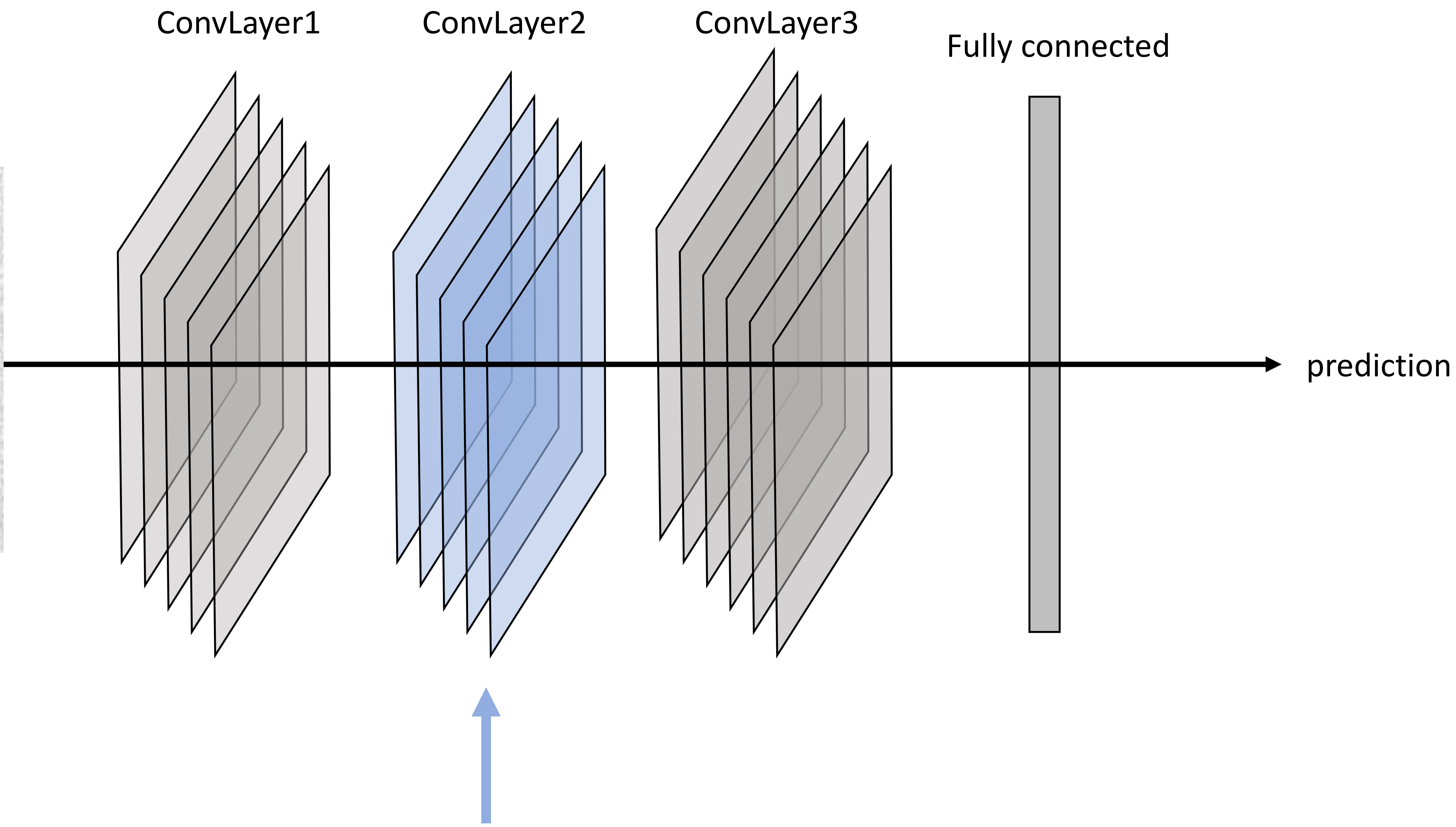


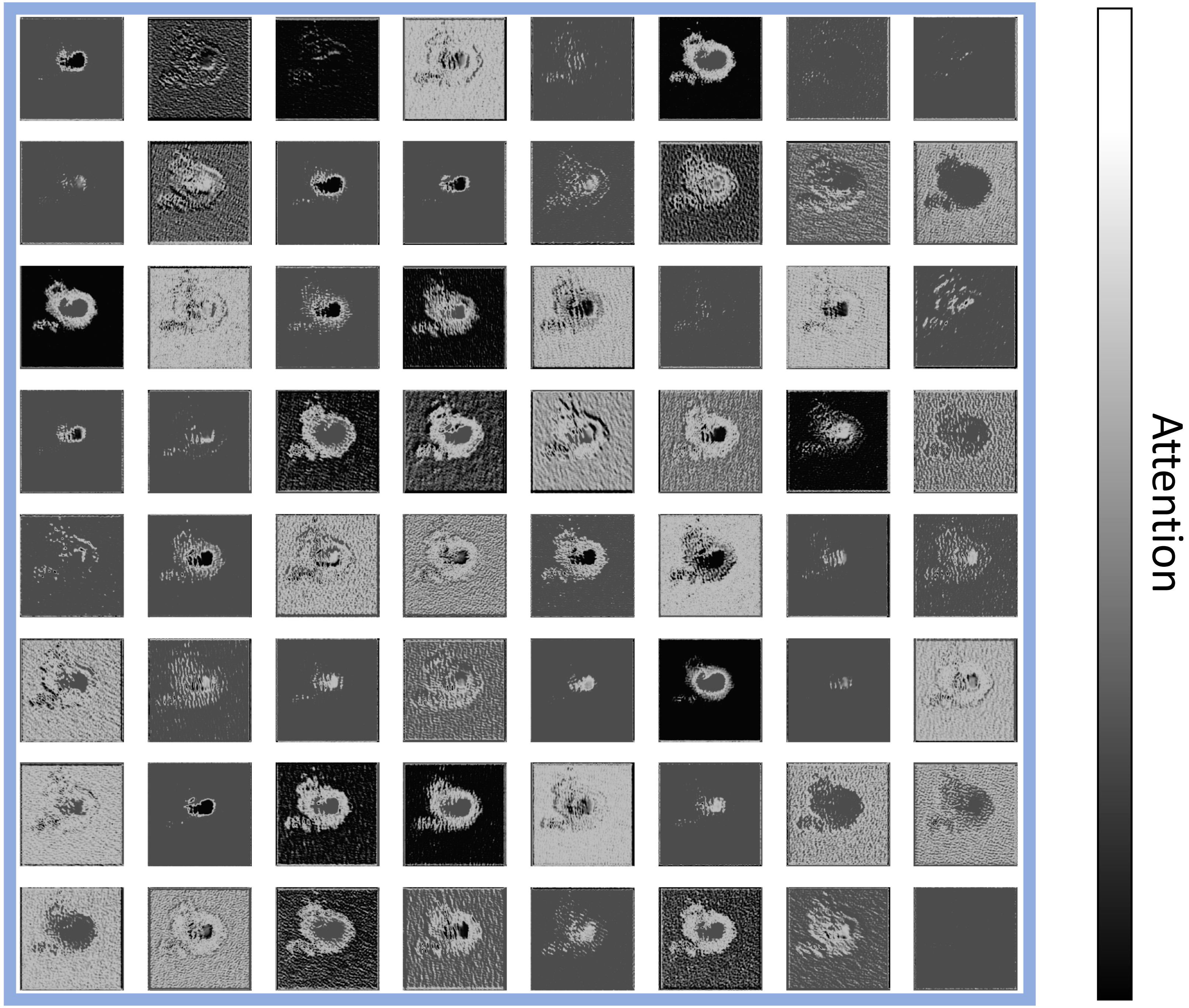

ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction
NASA/SDO

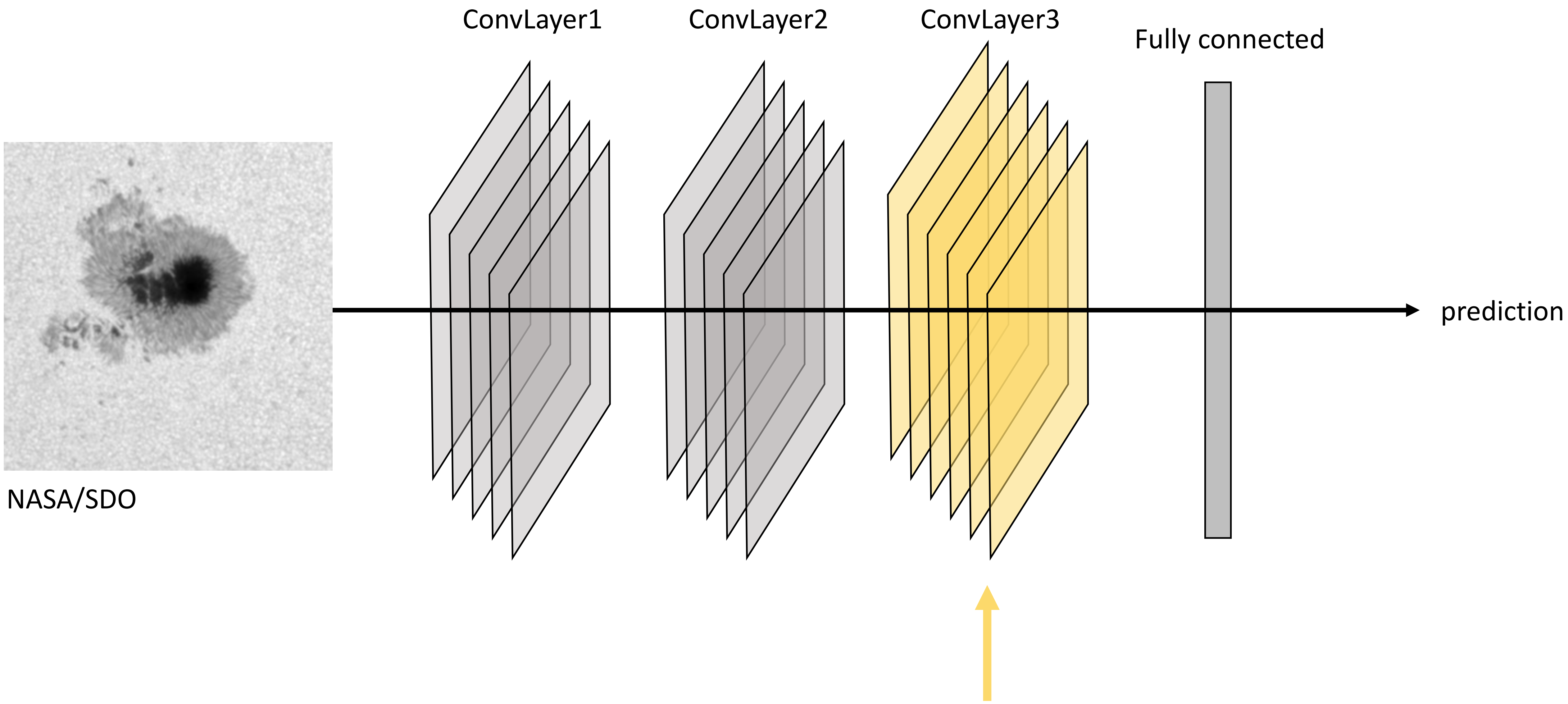

Penumbral border
Umbra
Granules
Attention



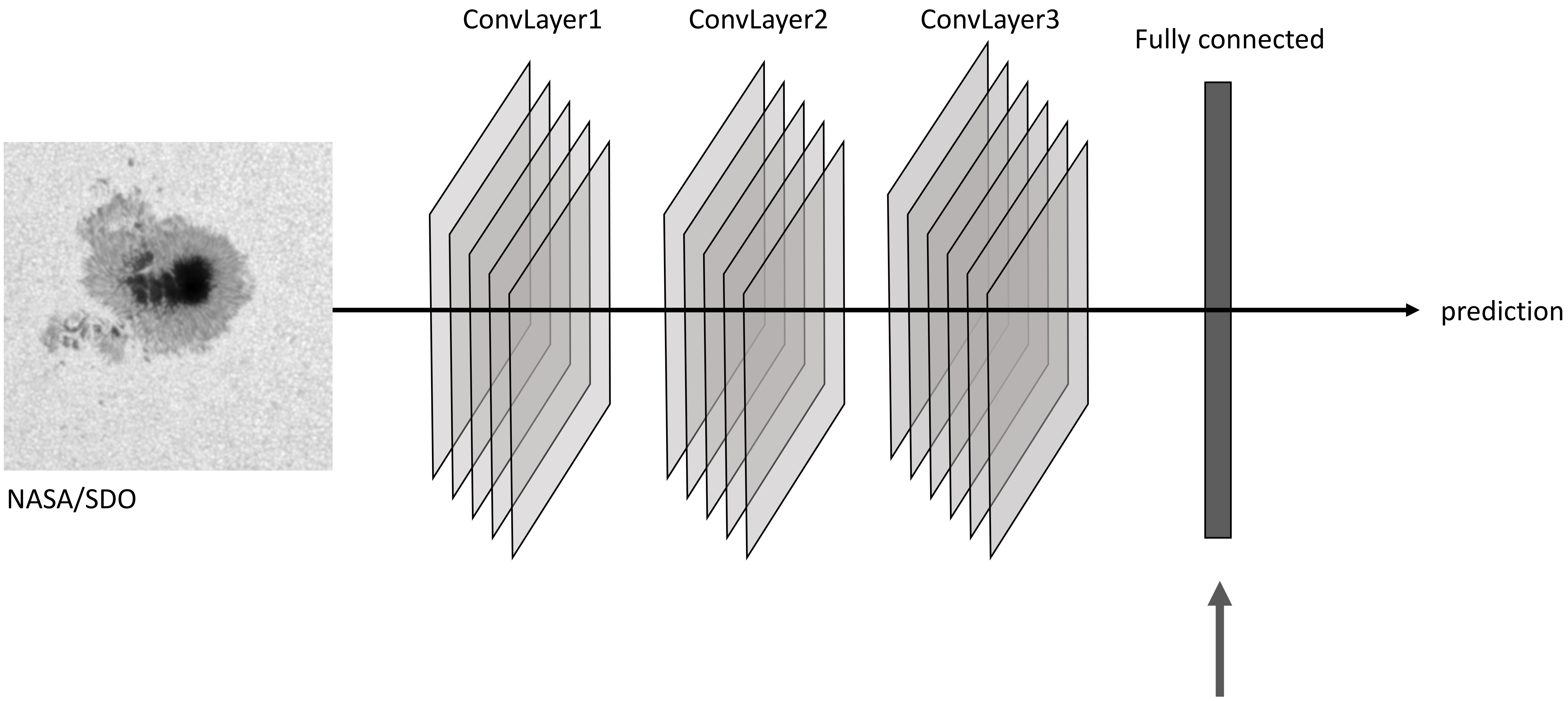
NASA/SDO


ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction

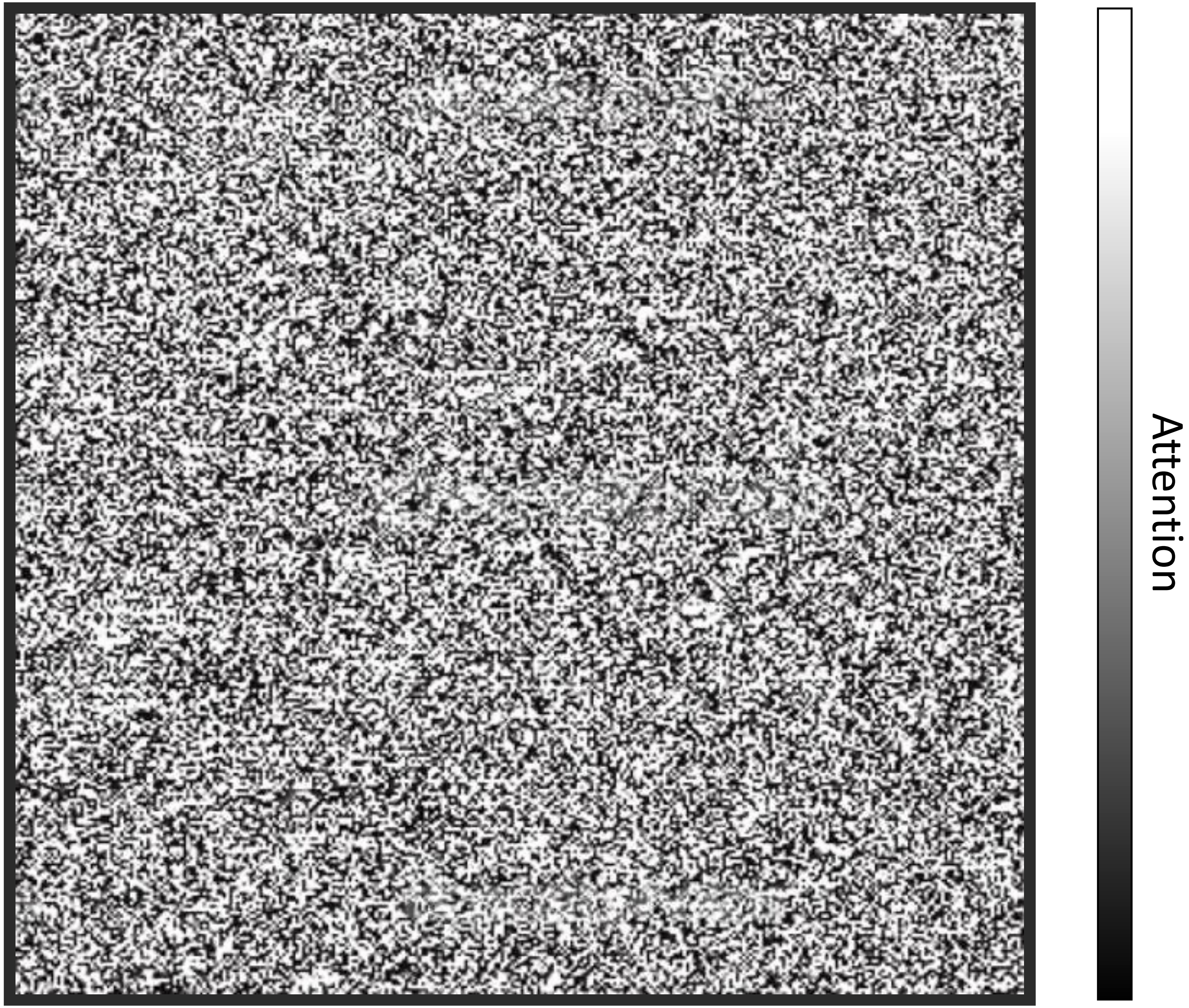

Attention

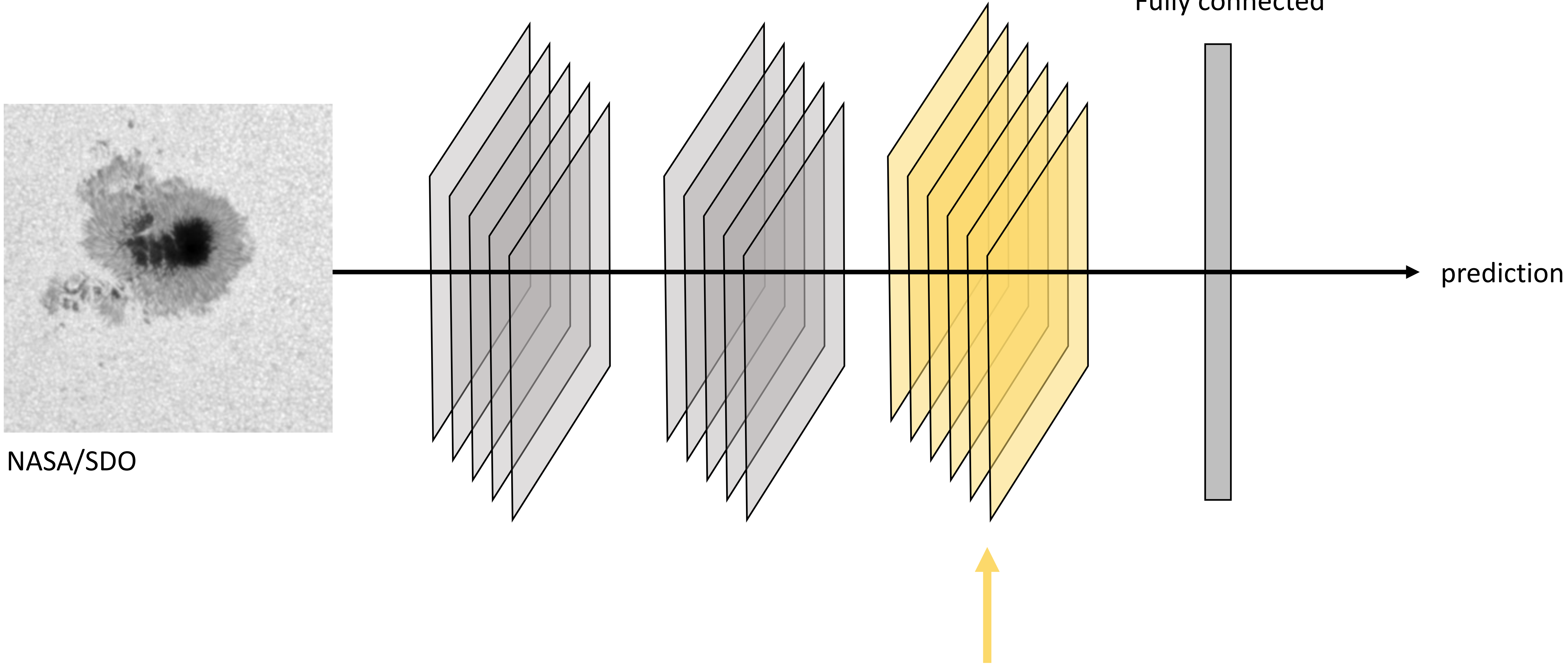

ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction
NASA/SDO


Attention


ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction
NASA/SDO

White Noise
Spatial coherence Is lost

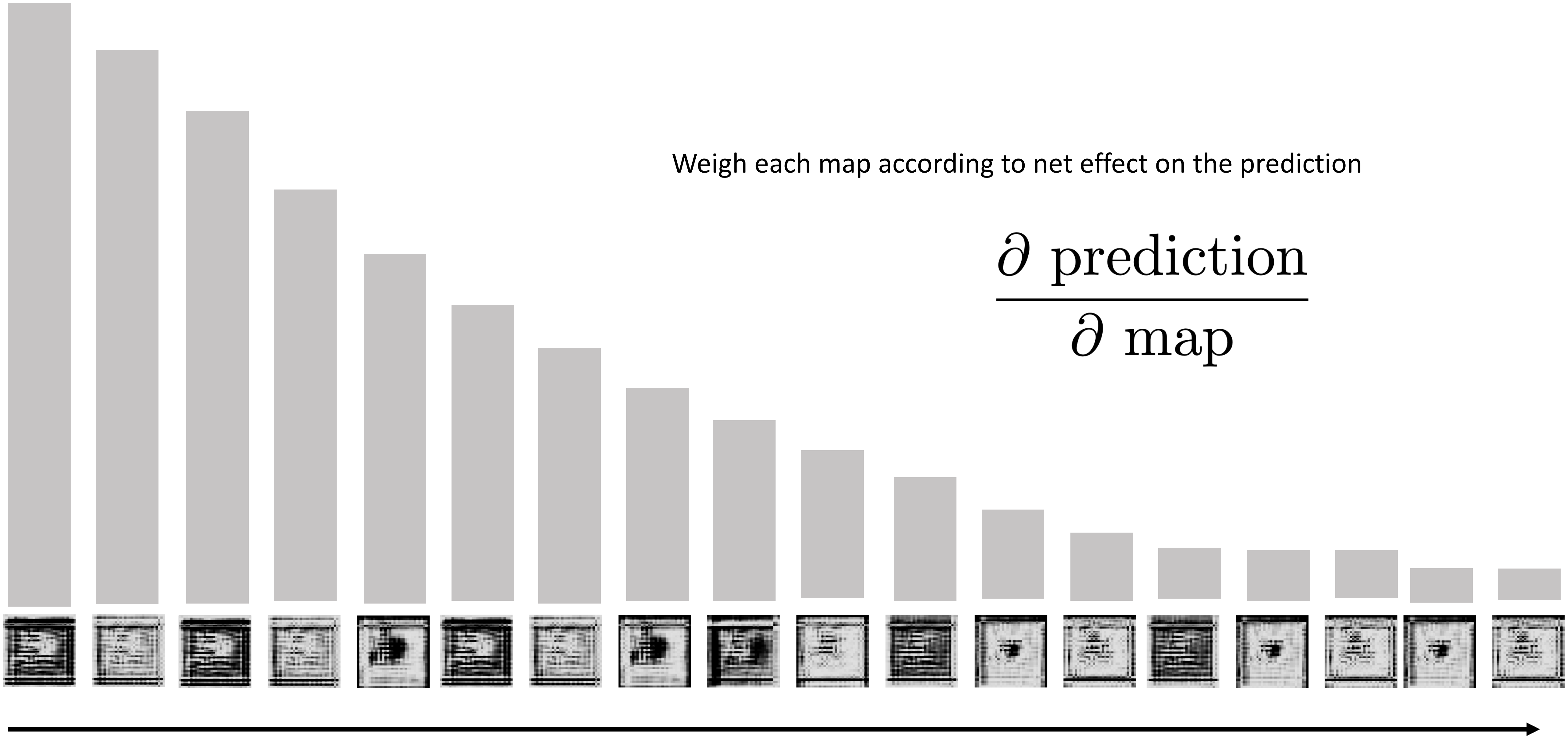



ConvLayer1
ConvLayer2
ConvLayer3
Fully connected
prediction

NASA/SDO

Wiggle each map and see how sensitive the output score is

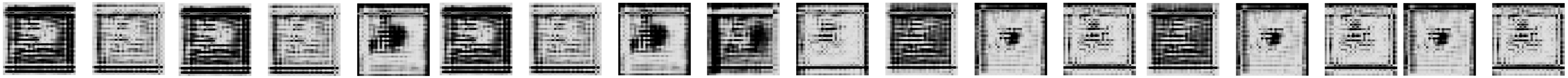
$$\frac{\partial \text{ prediction}}{\partial \text{ map}}$$

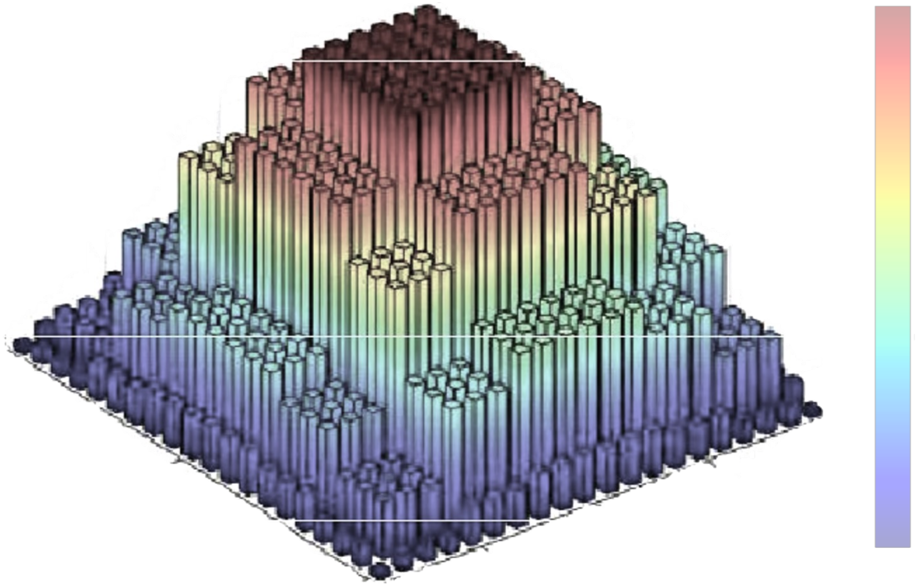


Weigh each map according to net effect on the prediction

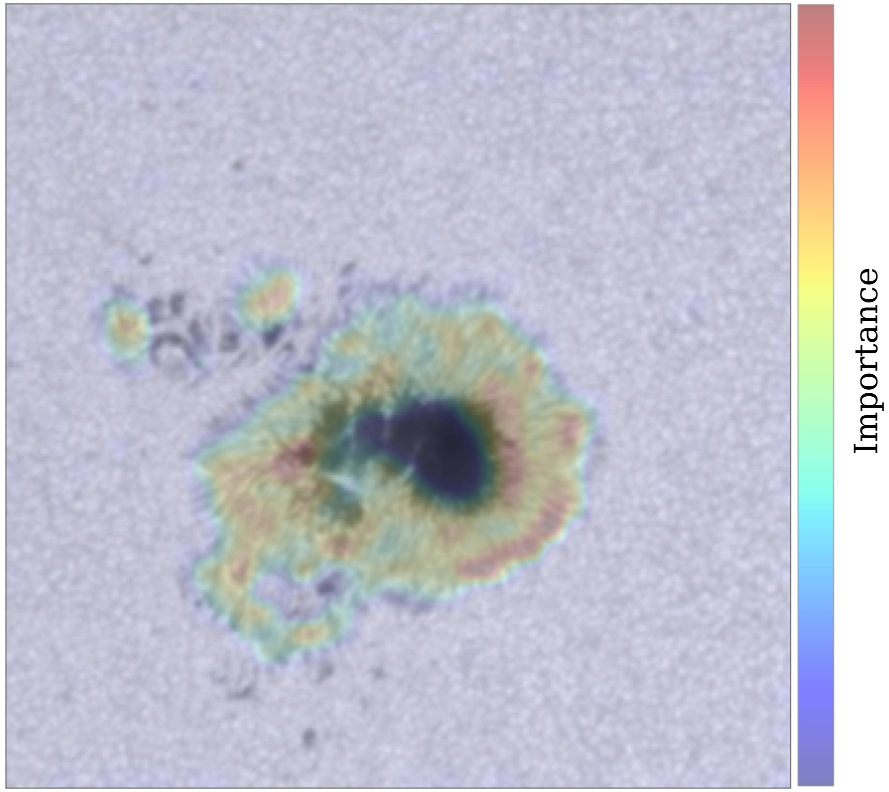
$$\frac{\partial \text{ prediction}}{\partial \text{ map}}$$

$$\sum_{\mathrm{i}} \lambda_i \mathrm{map}_i$$





The height (z) dimension is pixel importance



Grad-CAM highlights the penumbra, since this is what it searches for when classifying an image as a Sunspot

Blackbox model

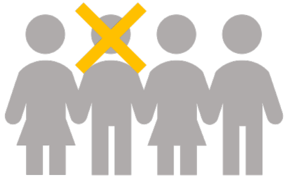
Input datapoint

$$\phi_i(f,x) = \sum_{z' \subseteq x'} \underbrace{\frac{|z'|!(M-|z'|-1)!}{M!}}_{\text{Weighting}} \underbrace{[f_x(z') - f_x(z' \setminus i)]}_{\text{Contribution}}$$

Shapley value for feature i

Subset

Simplified data input



F/ Blackbox = Market→ = Deep learning model

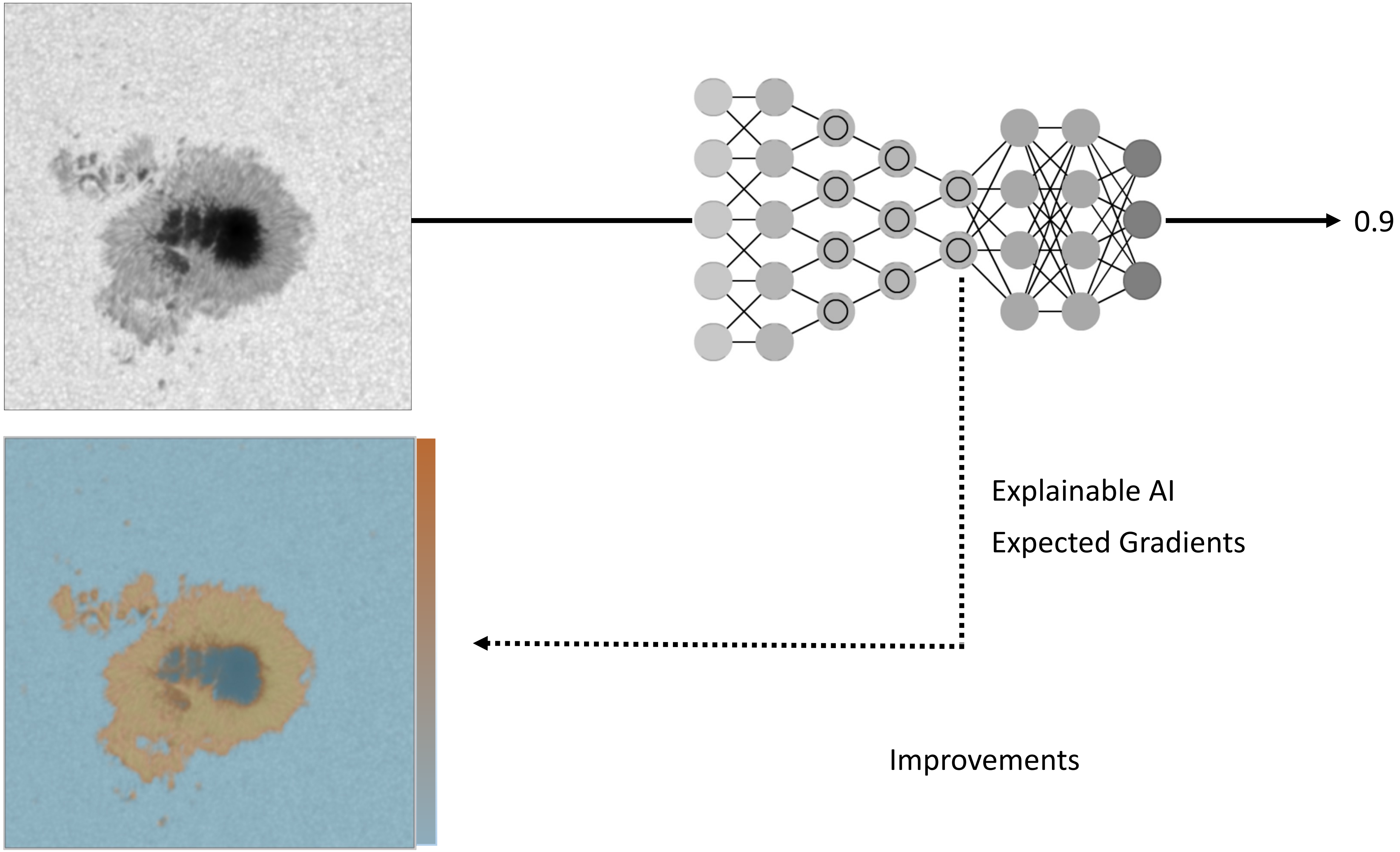
X = Coalition → image pixels

F(x) = pot of gold → prediction score [0,1]

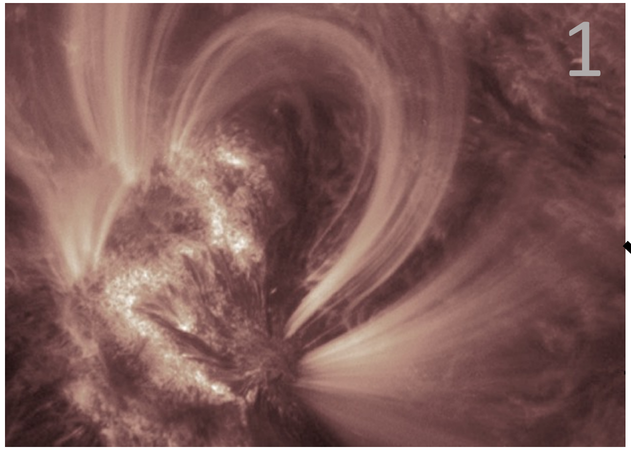
Subset……?

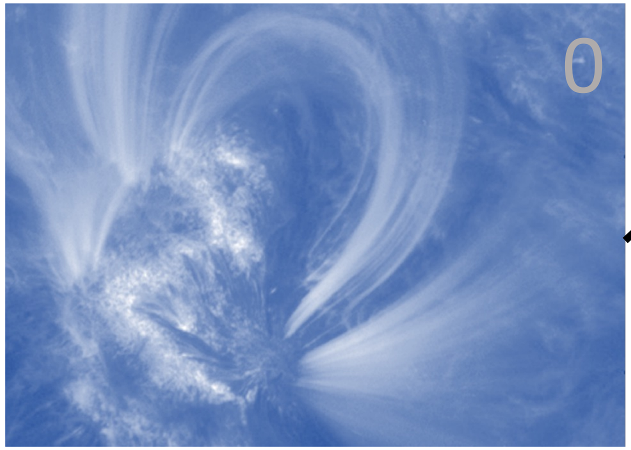
Zeros?, Gaussian?, Actual data?

$$\phi_\lambda|_{\mathrm{EG}} = \int_{x'} \Big((x_\lambda - x'_\lambda) \times \times \int_{\alpha=0}^{1} \frac{\delta \mathcal{F}_\Theta(x' + \alpha(x - x'))}{\delta x_\lambda} d\alpha\Big) p_D(x')\, dx' \simeq \underset{x' \sim D, \alpha \sim U(0,1)}{\mathbb{E}} \left[(x_\lambda - x'_\lambda) \frac{\delta \mathcal{F}_\Theta(x' + \alpha \times (x - x'))}{\delta x_\lambda} \right],$$

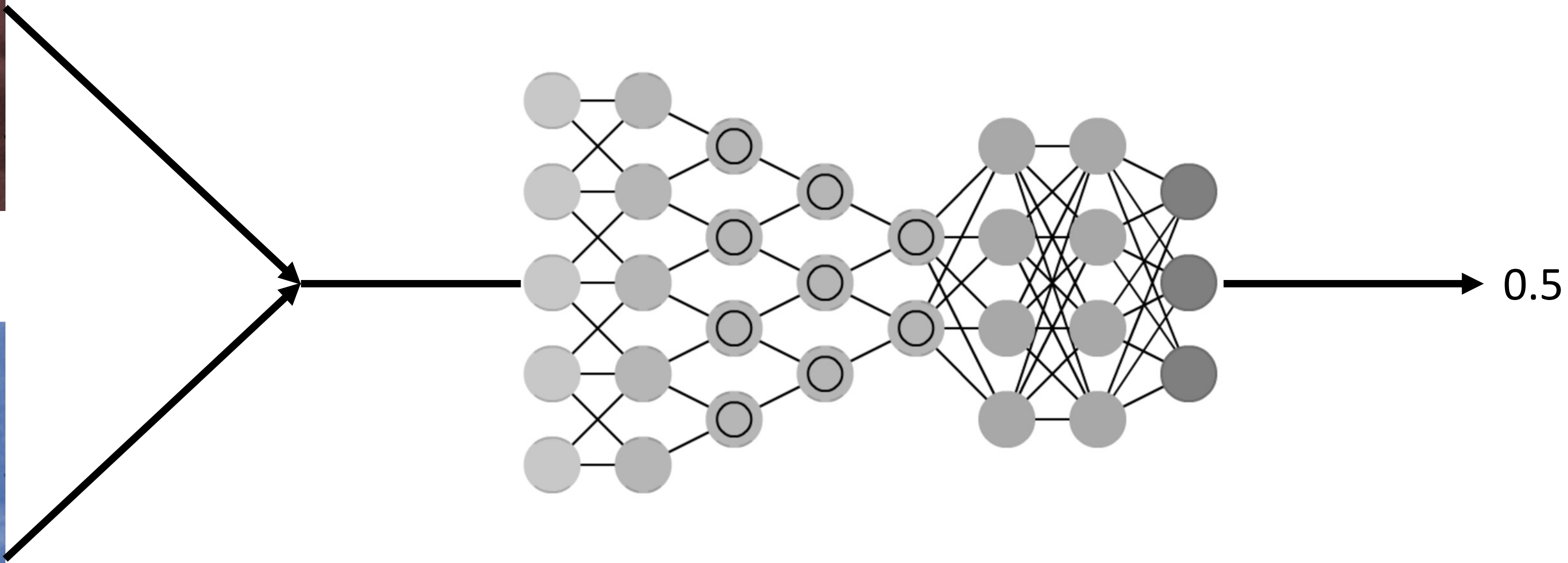

0.9
Explainable AI
Expected Gradients
Improvements

Spectra from pre-flare



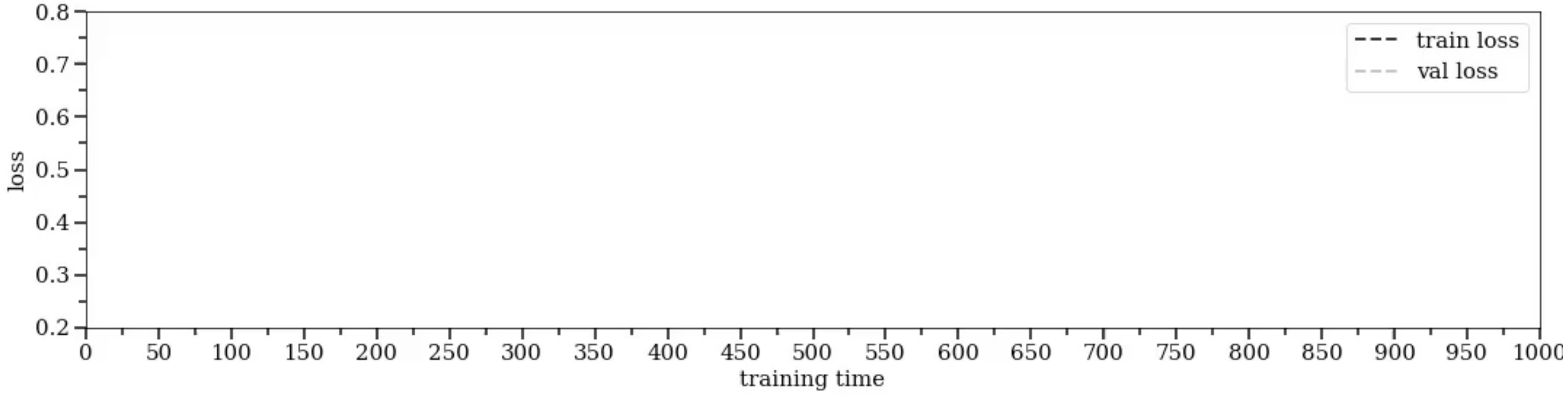


Spectra from active region

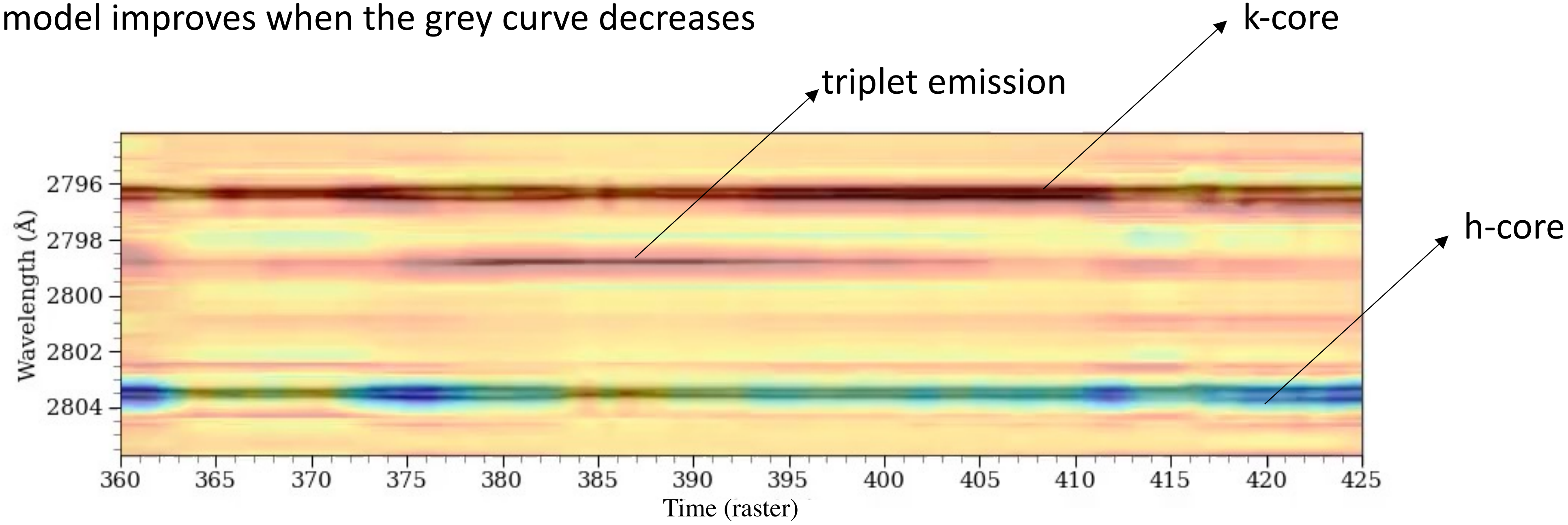


Once trained, the network should identify what features are strongly associated with precursory flare activity.

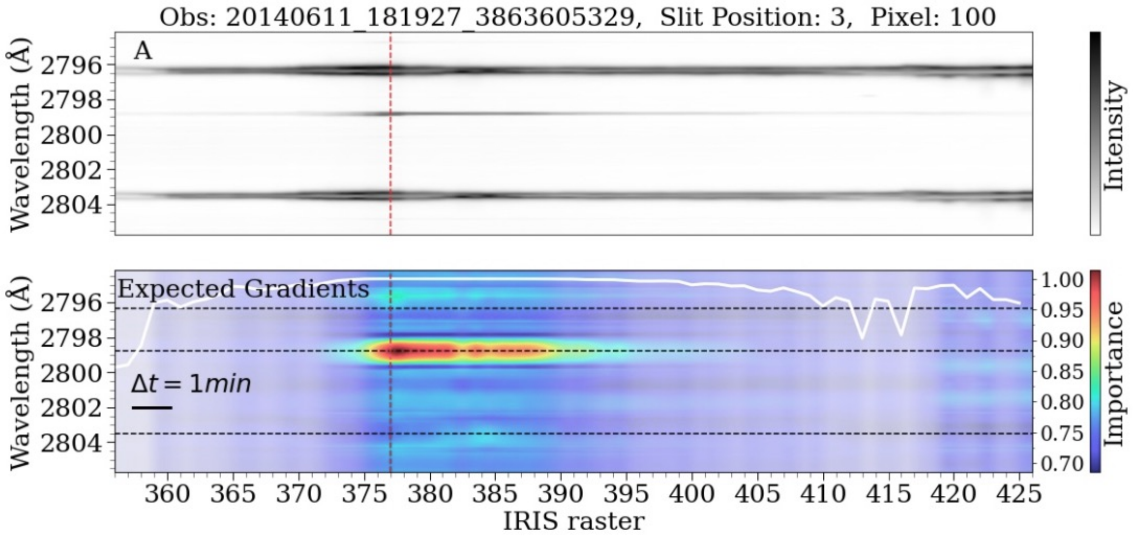
NASA/SDO



The model improves when the grey curve decreases

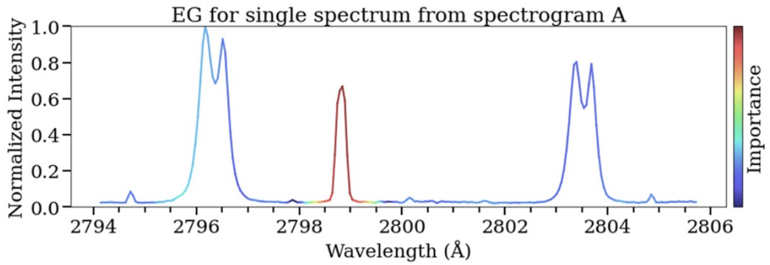


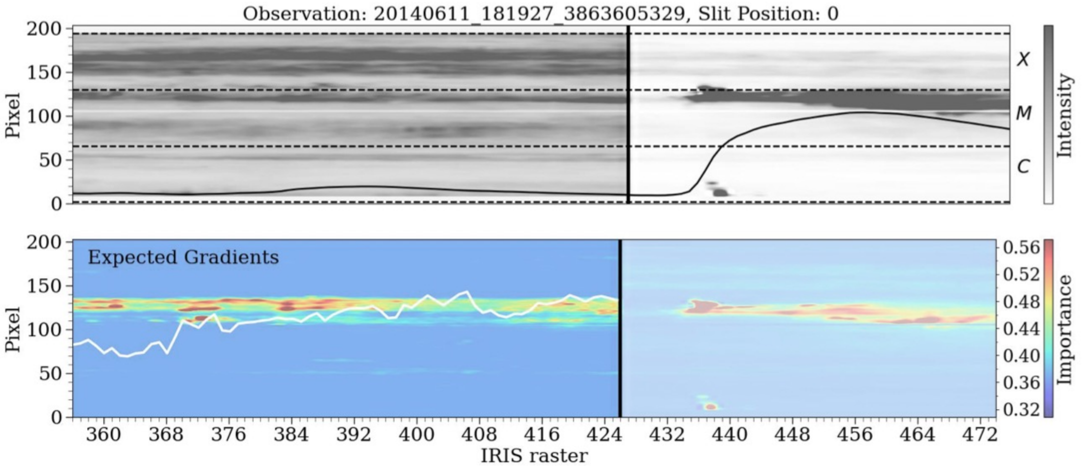
The network learns that triplet emission is important for deciding whether an active region will flare or not



B. Panos, L. Kleint, J. Zbinden A&A 671, A73 (2023)

Predicts triplet emission is nb for flare forecasting



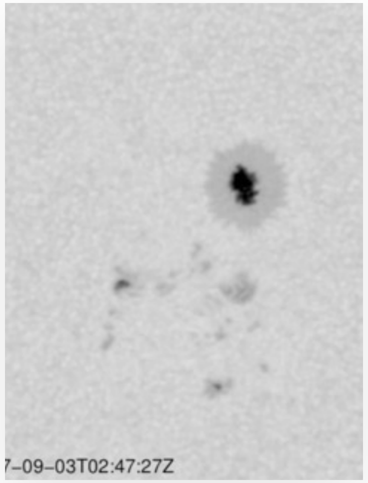


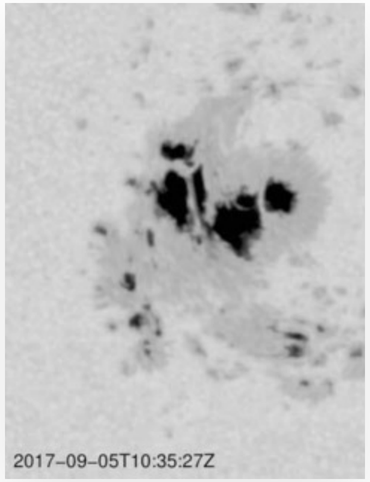
Model places attention where max future flare UV emission is seen later, even if more intense somewhere else

Modeling Sunspot evolution with transformer networks (Janis Witmer)

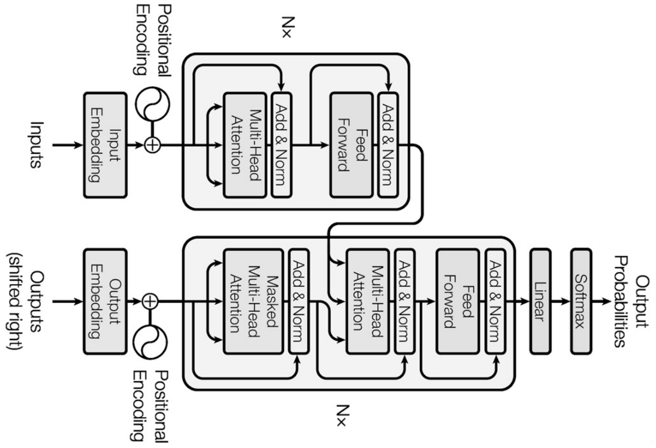
The idea

NASA/SDO



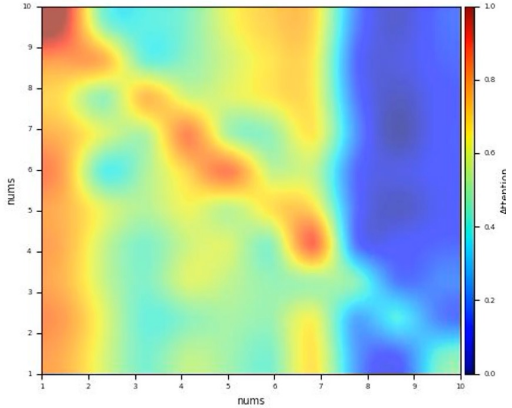


The model (ChatGPT)

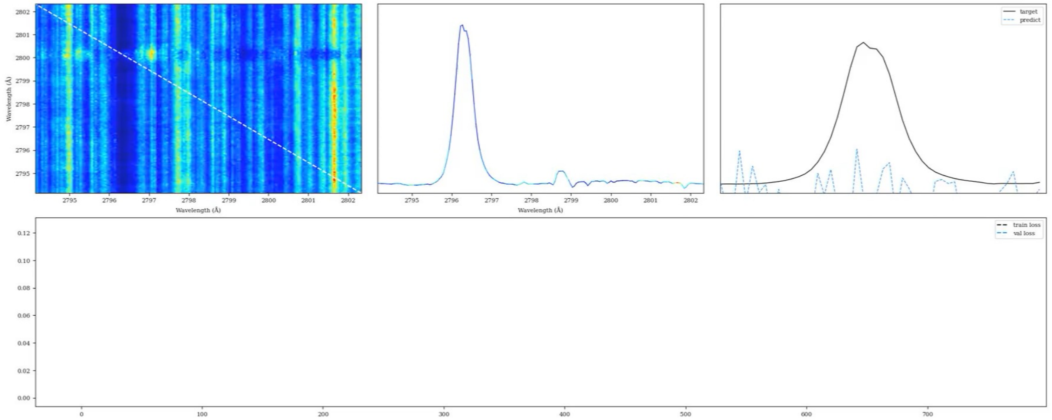


A. Vaswani et.al neurips (2017)

Reverse dataset



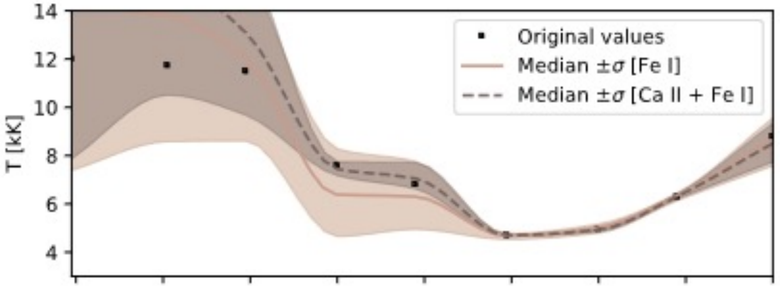
Spectral dataset



The tests

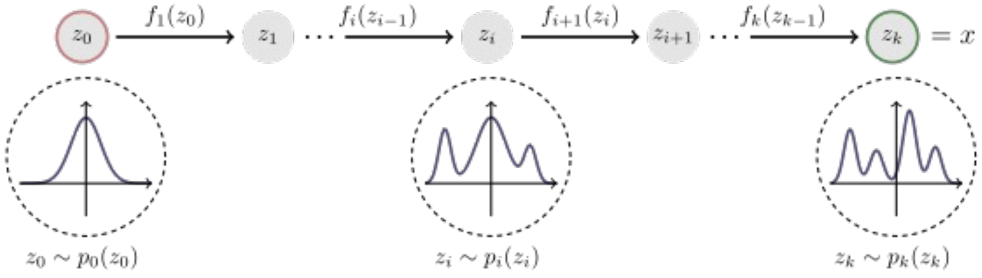
Normalizing flows for the Swedish Solar Telescope (Daniel Zahnd)

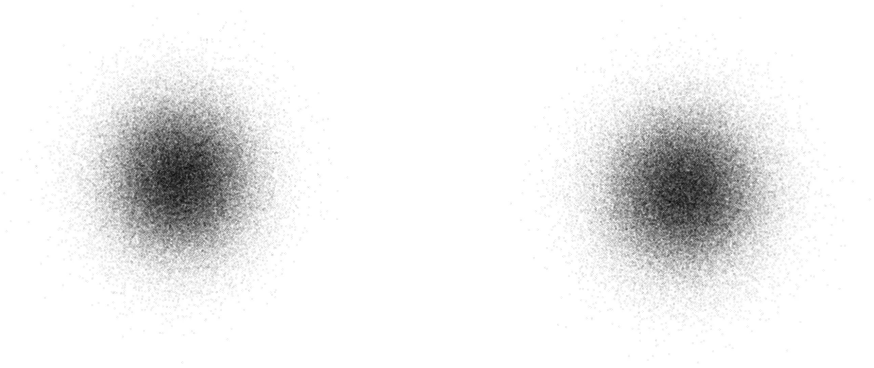
The idea

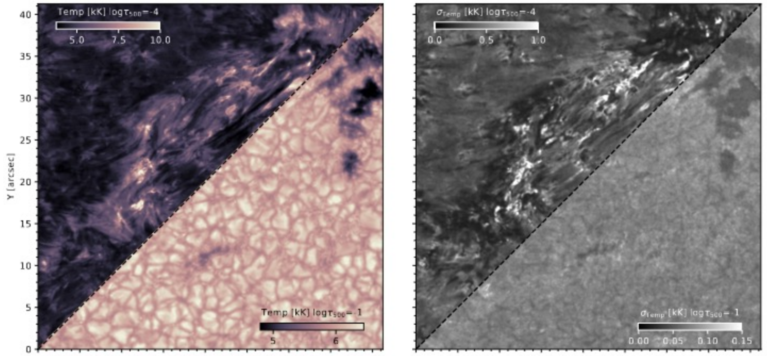


D. Baso et al A&A (2022)

The model (Normalizing flows)







D. Baso et al A&A (2022)

UNIVERSITY OF BERN

Thank you for your attention