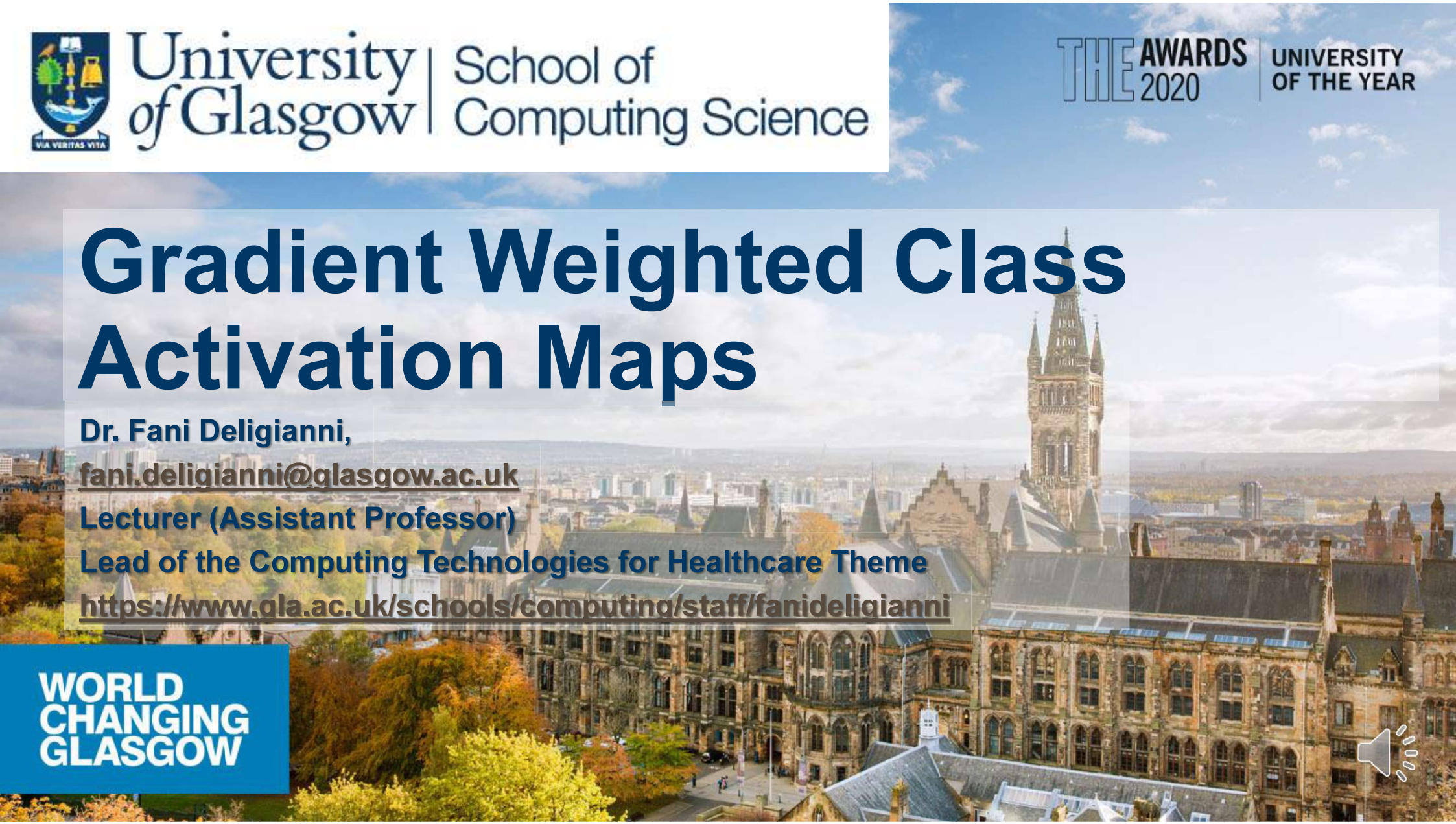


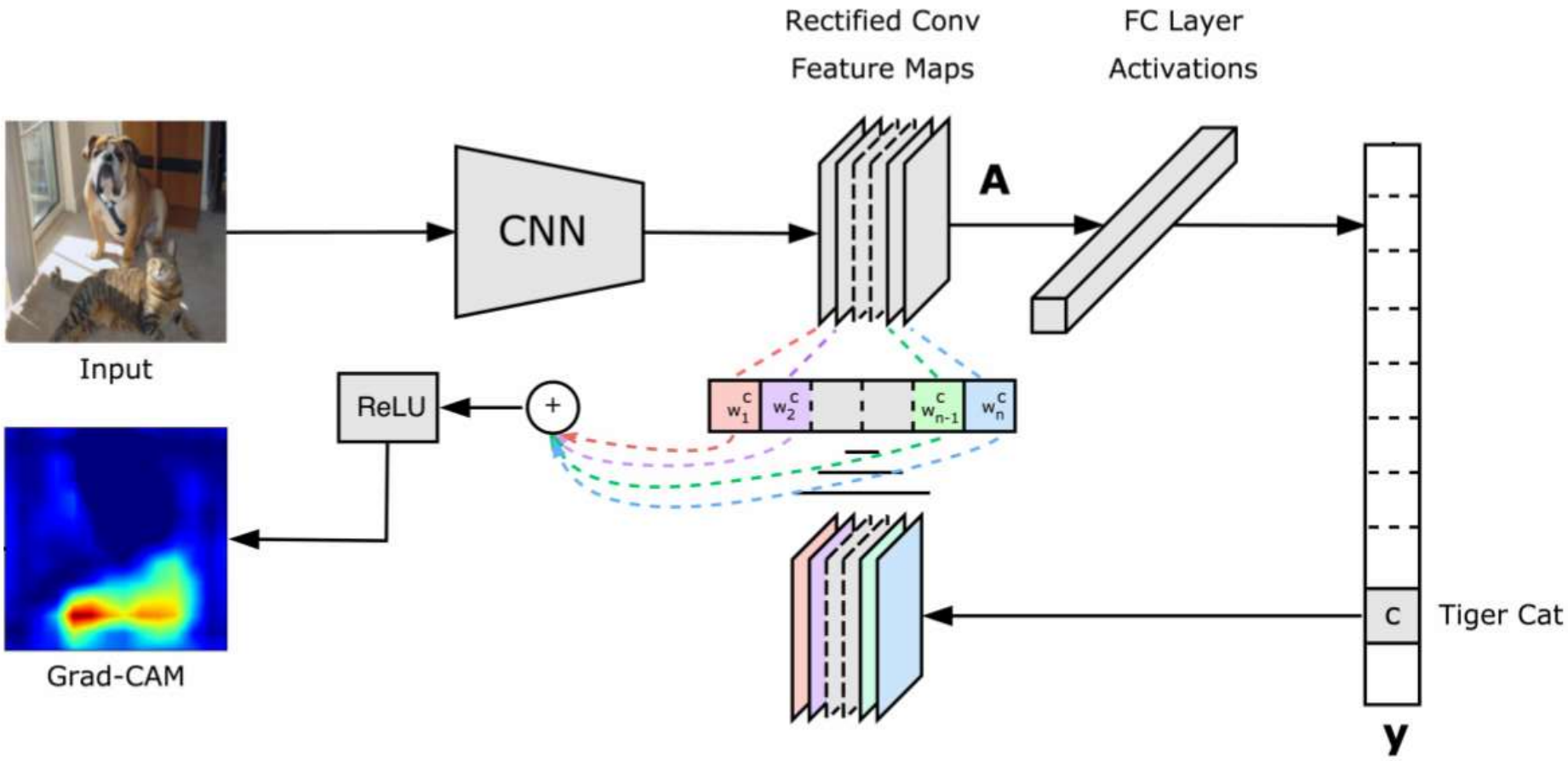
Gradient Weighted Class Activation Maps

Dr. Fani Deligianni,
fani.deligianni@glasgow.ac.uk
Lecturer (Assistant Professor)
Lead of the Computing Technologies for Healthcare Theme
https://www.gla.ac.uk/schools/computing/staff/fanideligianni

Model-Specific Approaches

- Guided Backpropagation
- Class Activations Maps
- **Gradient Weighted Class Activation Maps (GRAD-CAM)**
- **Guided Grad-CAM**

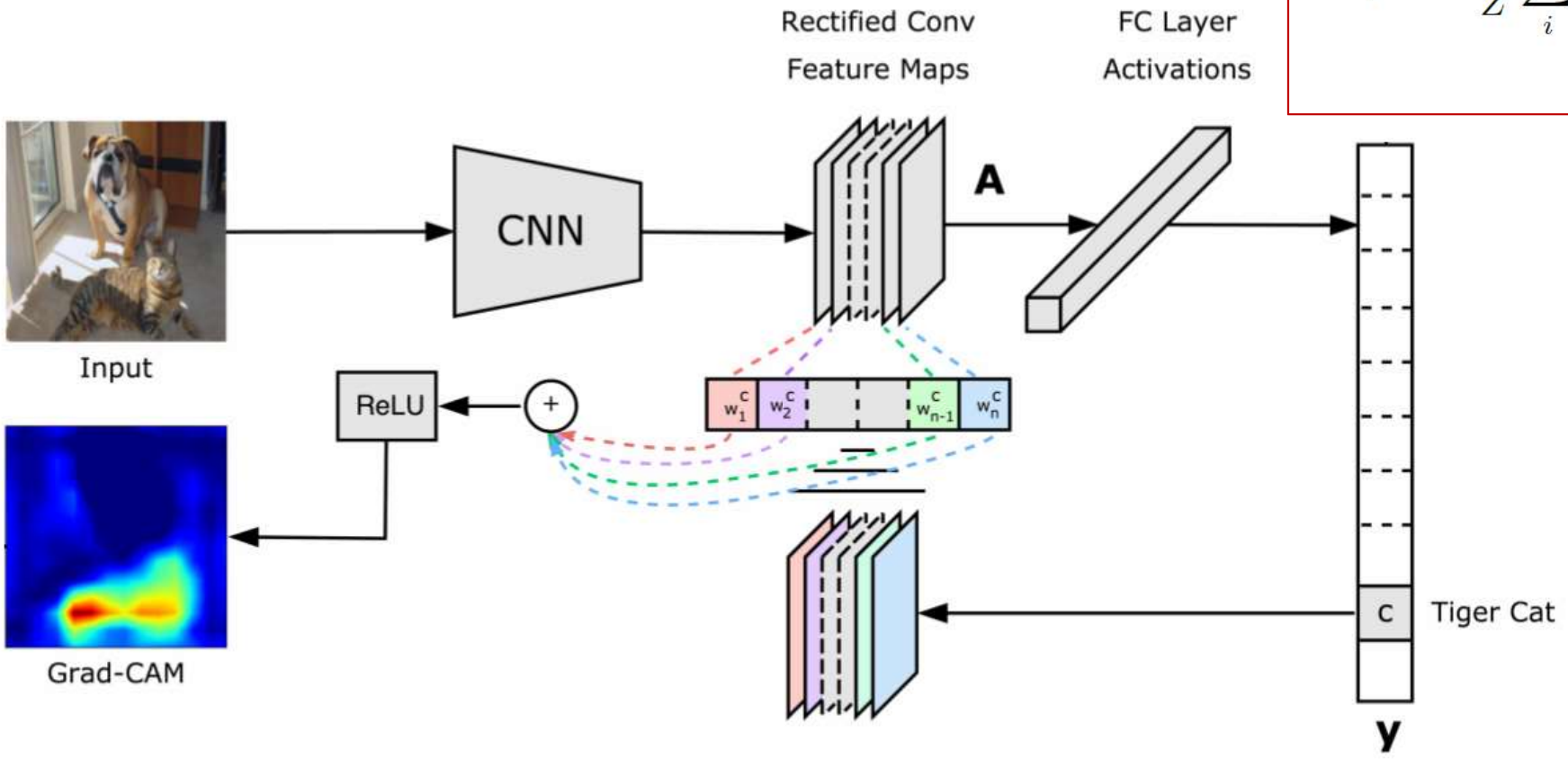
Gradient Weighted Class Activation Maps



Selvaraju et al. 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', IJCV, 2021

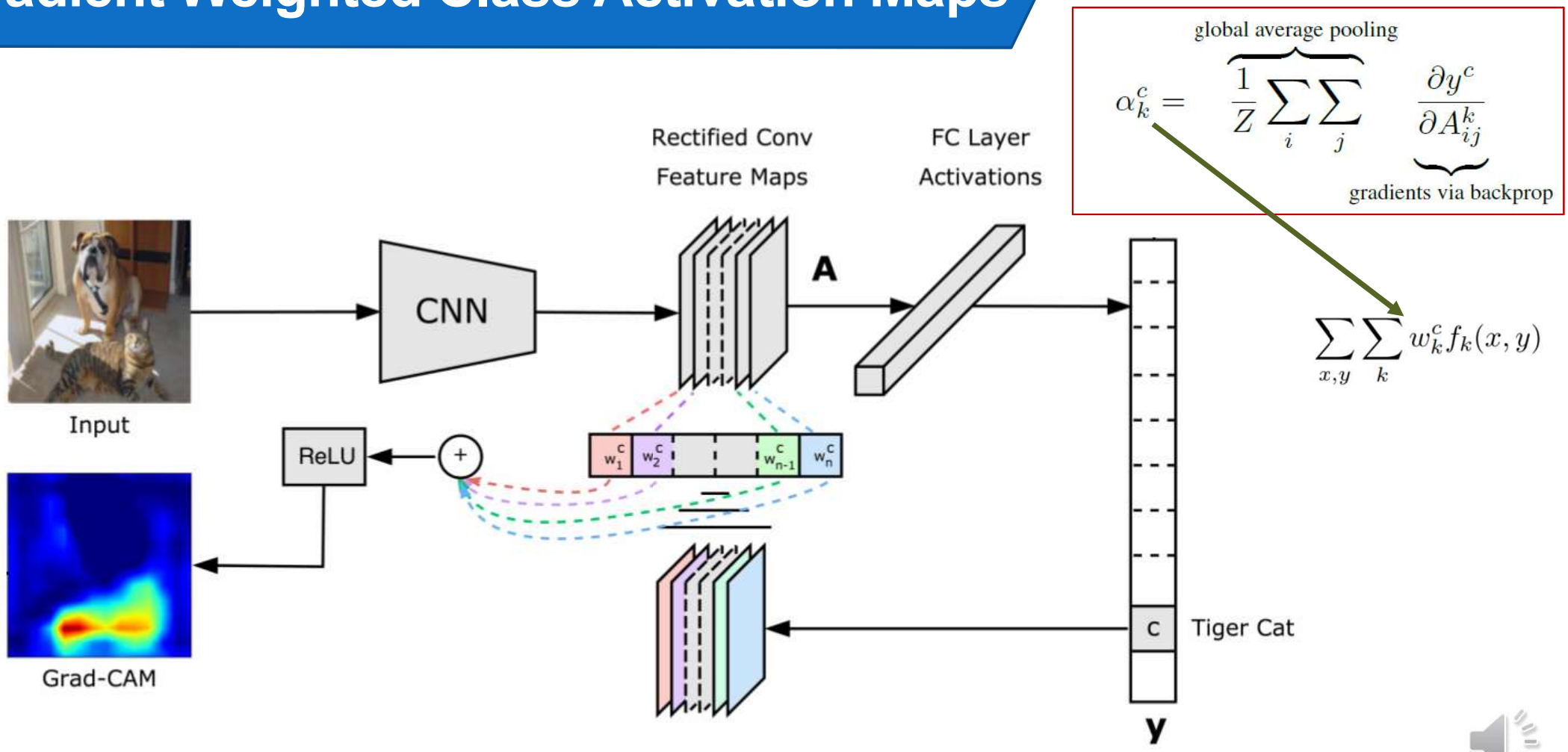
Gradient Weighted Class Activation Maps

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$



Selvaraju et al. 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', IJCV, 2021

Gradient Weighted Class Activation Maps

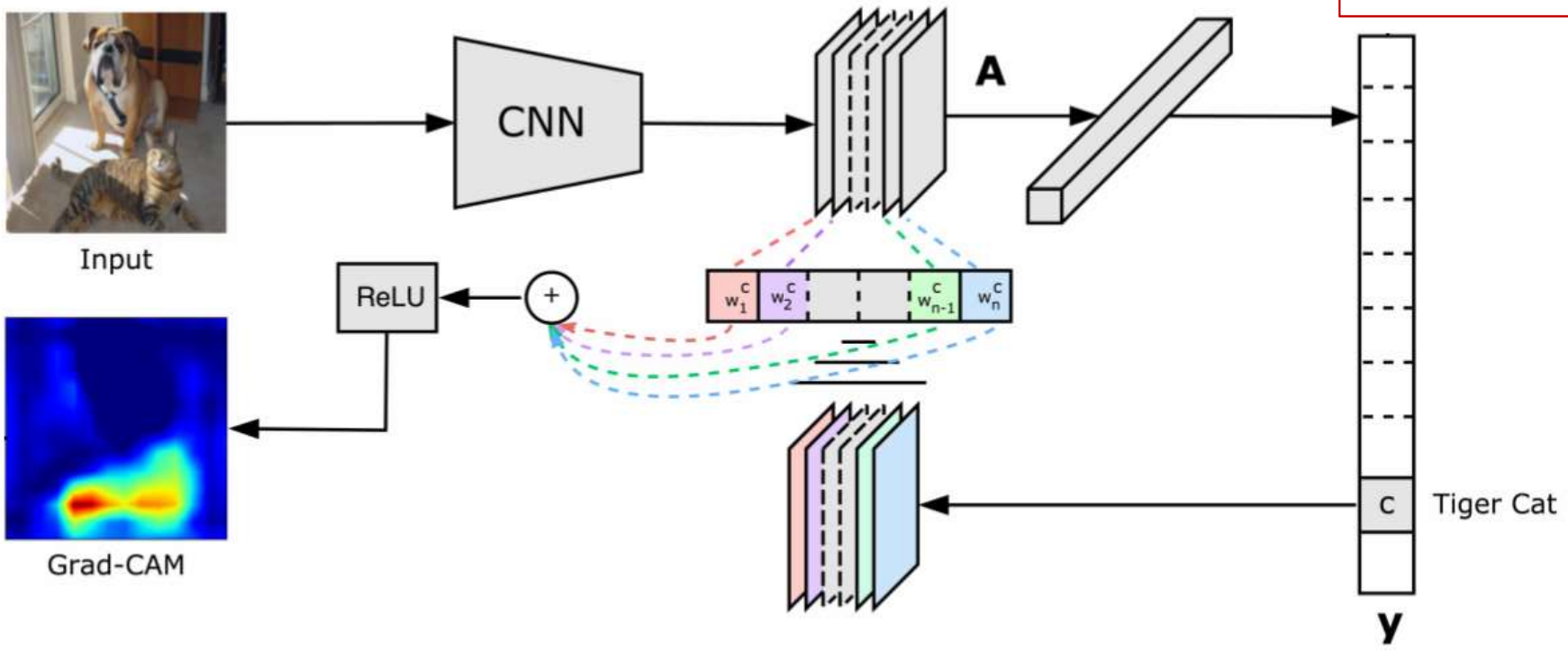


Selvaraju et al. 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', IJCV, 2021

Gradient Weighted Class Activation Maps

$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i\sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

$$L^c_{\text{Grad-CAM}} = ReLU\underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}}$$

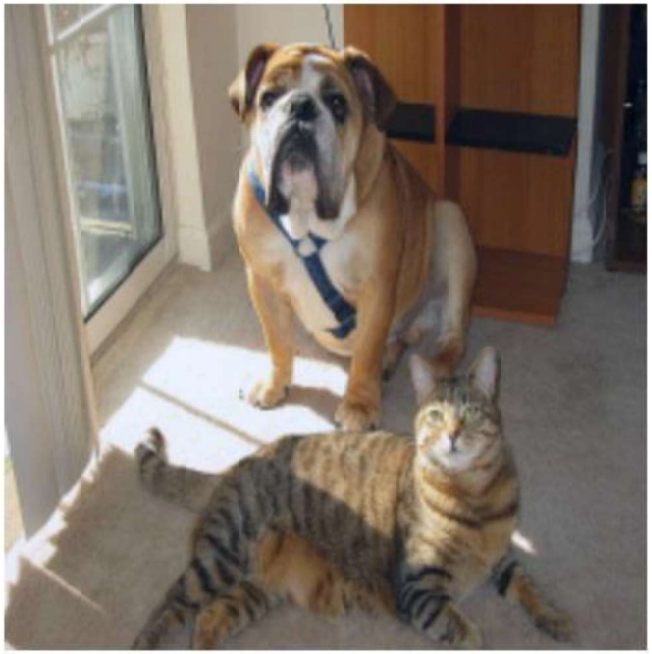


Selvaraju et al. 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', IJCV, 2021

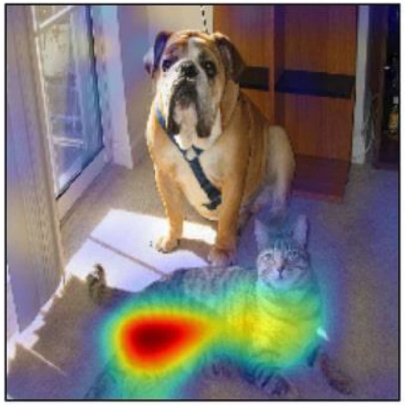
GRAD-CAM over CAM

- **Grad-CAM is a generalization of CAM**
- **Grad-CAM provides visualizations for a large family of CNNs:**
  - CNNs with fully connected layers
  - CNNs used for structured output
  - CNNs used in tasks with multi-modal inputs
- **Grad-CAM does not require retraining the deep neural network**

Grad-CAM



Guided Backpropagation ‘Cat’



Grad-CAM ‘Cat’

Guided Backpropagation ‘Dog’

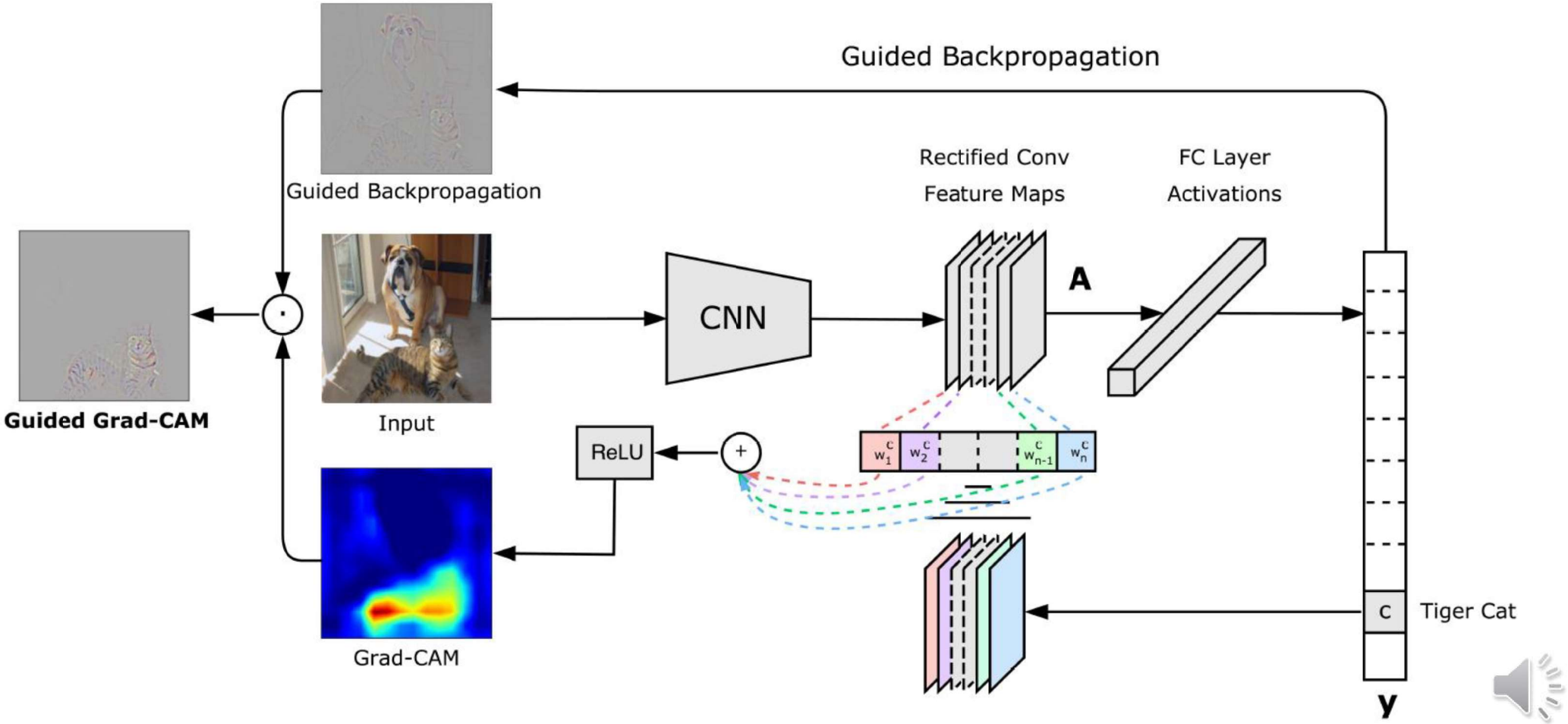
Grad-CAM ‘Dog’

Class-discriminative ✓

High-resolution ✗

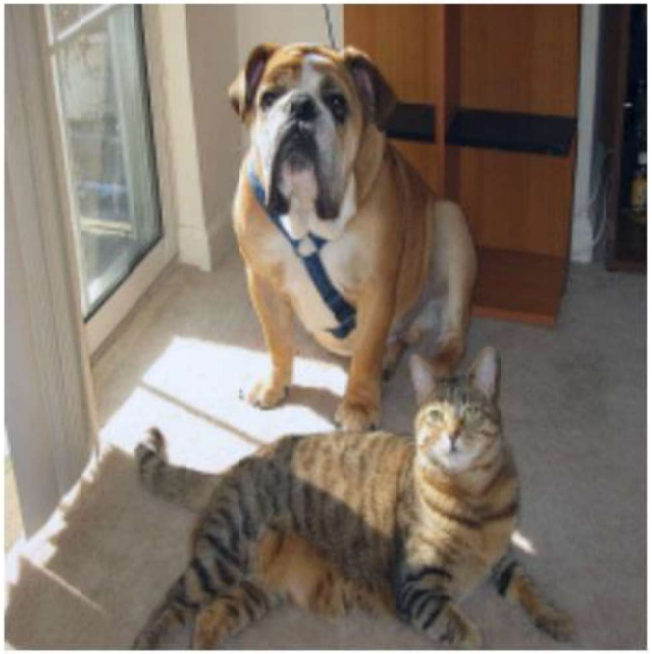
It can apply in more architectures ✓

Guided GRAD-CAM

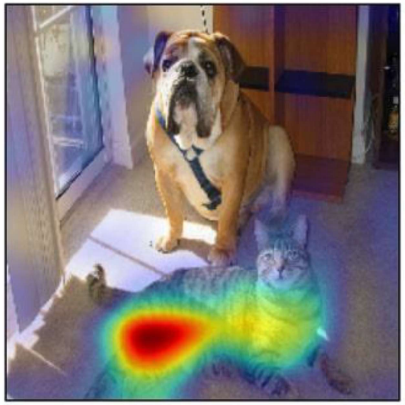


Selvaraju et al. 'Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization', IJCV, 2021

Guided Backpropagation



Guided Backpropagation ‘Cat’



Grad-CAM ‘Cat’

Guided Grad-CAM ‘Cat’

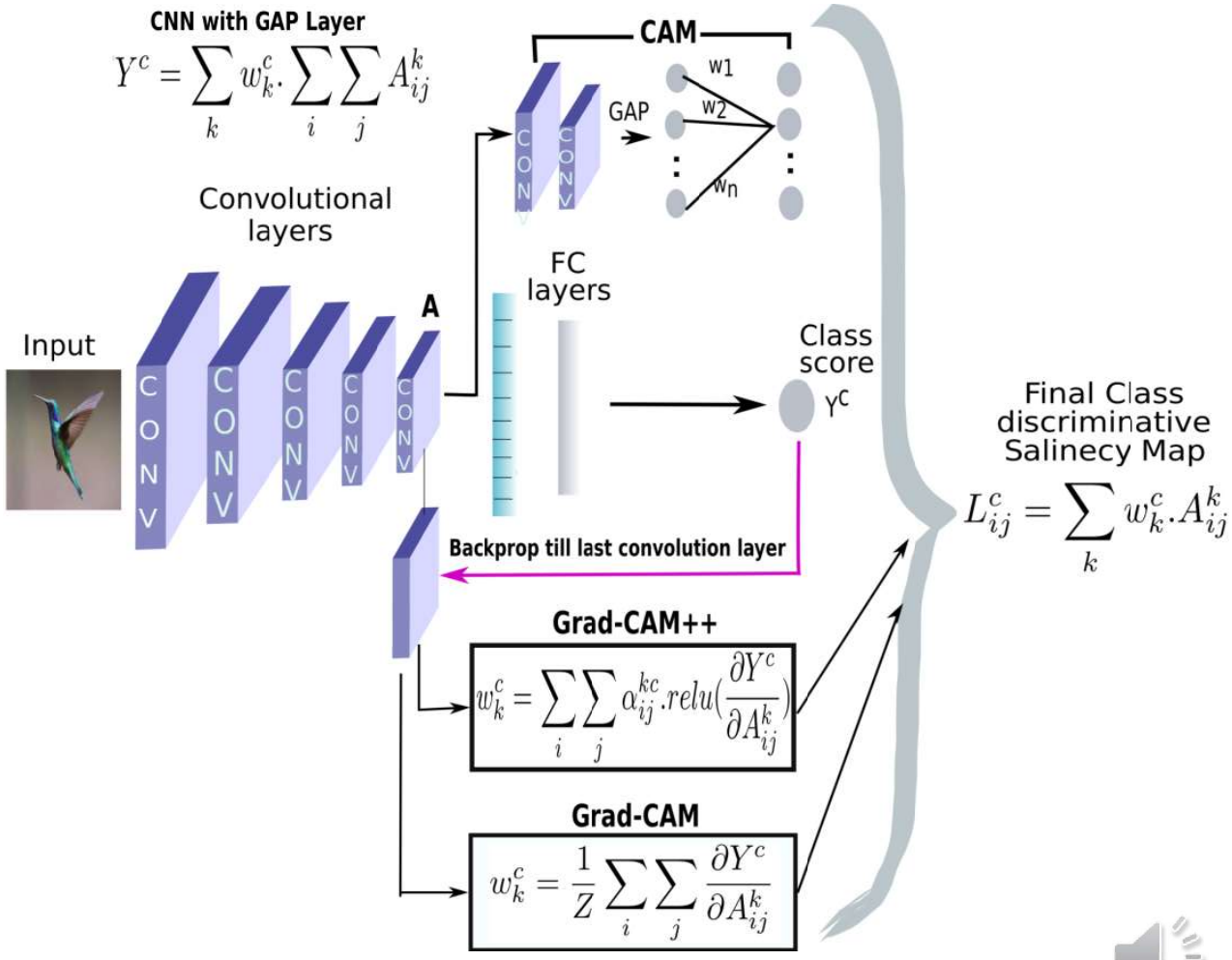
Guided Backpropagation ‘Dog’

Grad-CAM ‘Dog’

Guided Grad-CAM ‘Dog’

Grad-CAM++

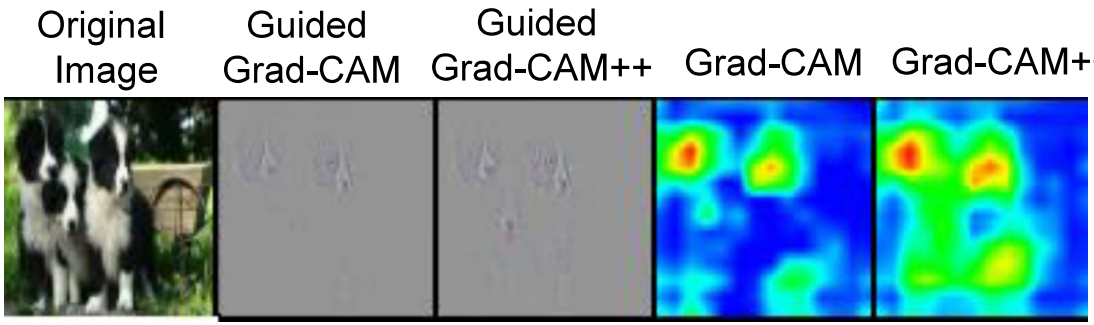
- It considers only the positive weights of the gradients
- It improves sensitivity and coverage of Grad-CAM
- Explaining multiple instances of an object in an image

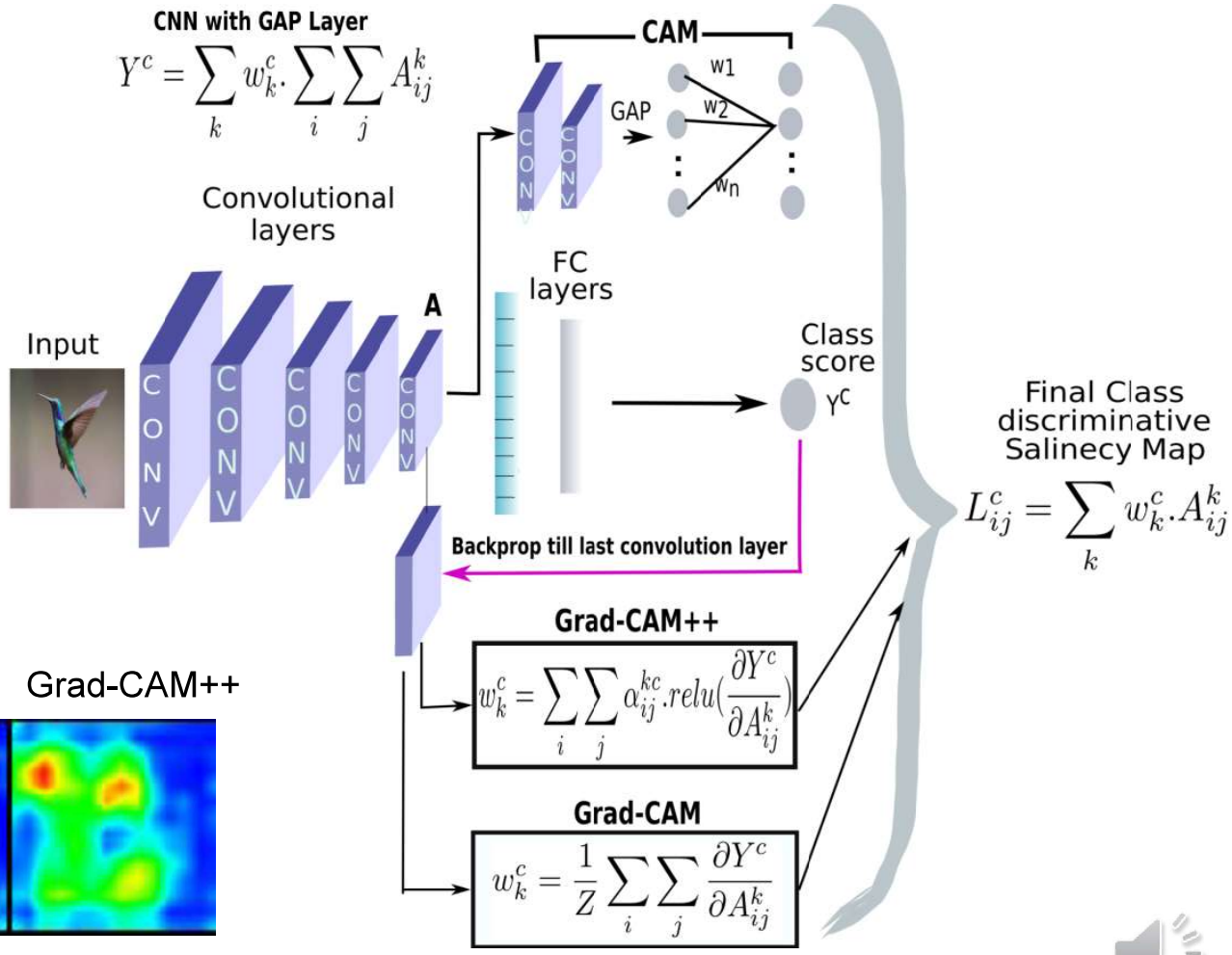


Chattopadhya et al. 'Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks', WACV, 2021

Grad-CAM++

- It considers only the positive weights of the gradients
- It improves sensitivity and coverage of Grad-CAM
- Explaining multiple instances of an object in an image





Chattopadhya et al. 'Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks', WACV, 2021

Summary

- Grad-CAM is a generation of Class Activation Mapping that allows more flexibility with relation to the network architectures
- Grad-CAM can be combined with Guided-Backpropagation to provide both class-discriminative and high-resolution visualization maps
- Several variants of Grad-CAM method have been also developed, such as Grad-CAM++

References

- Zhou et al. ‘Learning Deep Features for Discriminative Localization’, CVPR, 2016.
- Selvaraju et al. ‘Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization’, International Journal of Computer Vision, 2019.
- Chattopadhya et al. ‘Grad-CAM++: Improved Visual Explanations for Deep Convolutional Networks’, WACV, 2021