

# Supplemental Material

## Jitter-CAM: Improving the Spatial Resolution of CAM-Based Explanations

Thomas Hartley<sup>1</sup>  
hartleytw@cardiff.ac.uk

Kirill Sidorov<sup>1</sup>  
sidorovk@cardiff.ac.uk

Christopher Willis<sup>2</sup>  
chris.willis@baesystems.com

David Marshall<sup>1</sup>  
marshallad@cardiff.ac.uk

<sup>1</sup> Cardiff University

<sup>2</sup> BAE Systems Applied Intelligence

---

## 1 Additional Examples

Here we present additional examples for all the models discussed in the main paper. All example images are taken from the ImageNet validation set and are classified correctly by the models. We show examples for the following models:

- Explanations created for ResNet50 are shown in Figure 1.
- Explanations created for DenseNet121 are shown in Figure 2.
- Explanations created for InceptionV3 are shown in Figure 3.

In these examples we see the same trends as noted in the paper. All of the previous CAM methods are both visually similar, and very coarse. For both ResNet50 and DenseNet121 we see that RISE produces more detailed explanations, although these are still coarser than Jitter-CAM. Of interest though are the RISE examples for InceptionV3. These explanations are much noisier using this architecture. This is likely due to increased size of the grid underpinning the RISE explanation prior to resizing. This is likely to require more passes than the other models in order to generate accurate explanations. However, a hyper-parameter search for RISE is out of the scope of this paper.

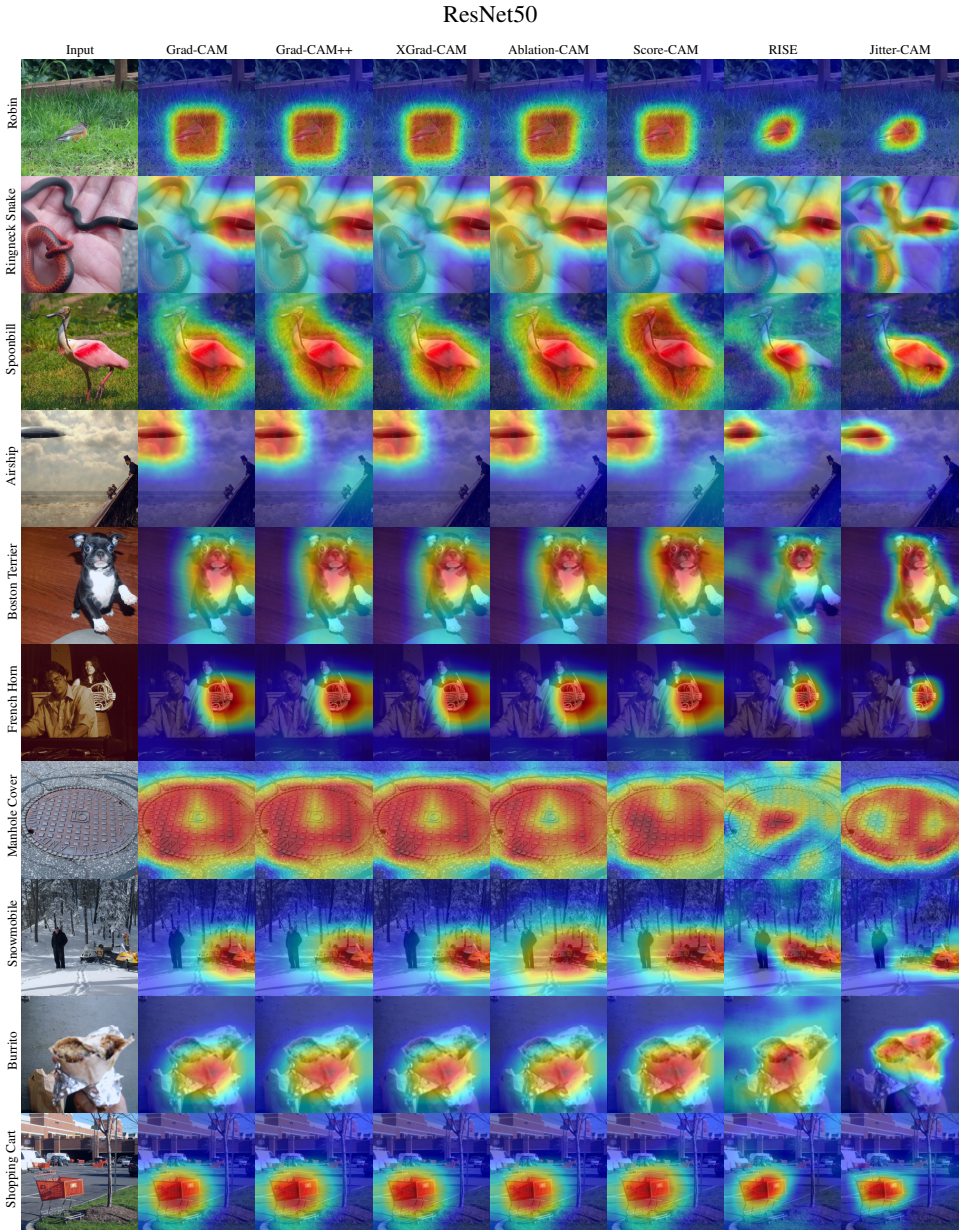


Figure 1: Additional examples for ResNet50 using ImageNet.

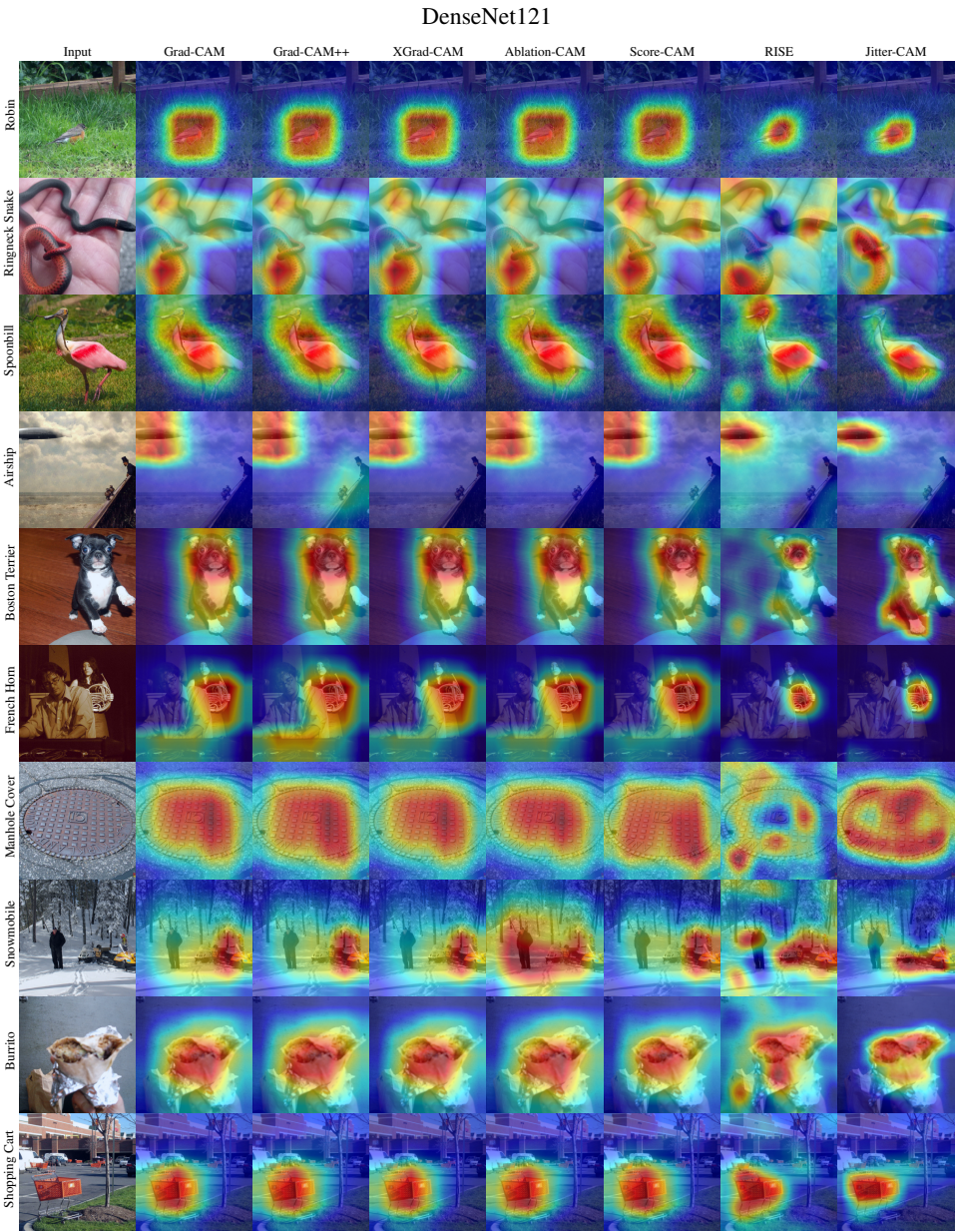


Figure 2: Additional examples for DenseNet121 using ImageNet



## InceptionV3

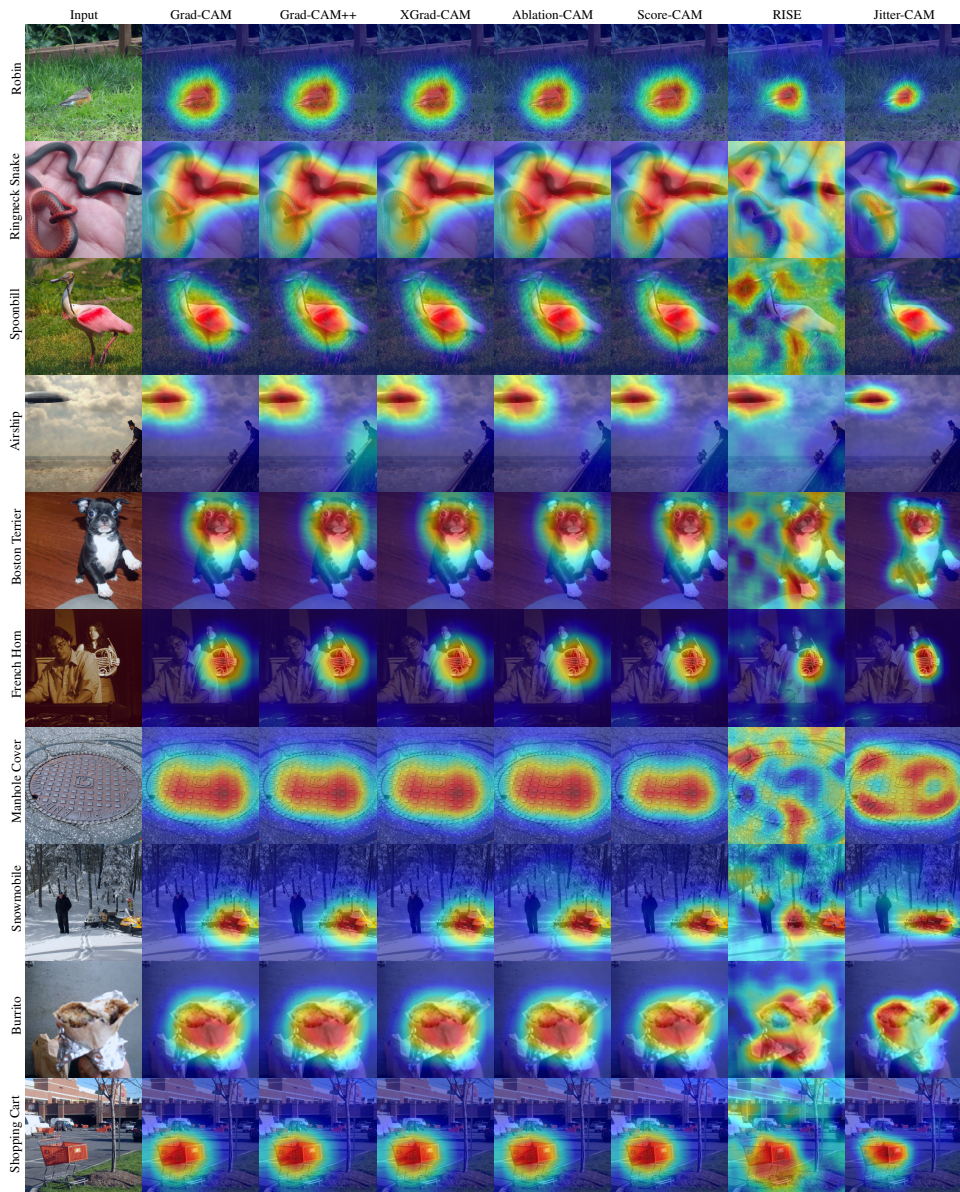


Figure 3: Additional examples for InceptionV3 using ImageNet.



## 2 5x5 -> 7x7 Examples

In this section we show visual examples of explanations resized from  $5 \times 5$  to  $7 \times 7$  using our Jitter-CAM technique. In Figure 4 we show the original image, the  $5 \times 5$  to  $7 \times 7$  Jitter-CAM explanation, and the original  $7 \times 7$  Grad-CAM explanation. To offer further insight we also show the Spearman score alongside the explanations. Included here is an example with a low Spearman score, despite this, similar regions are identified as being important.

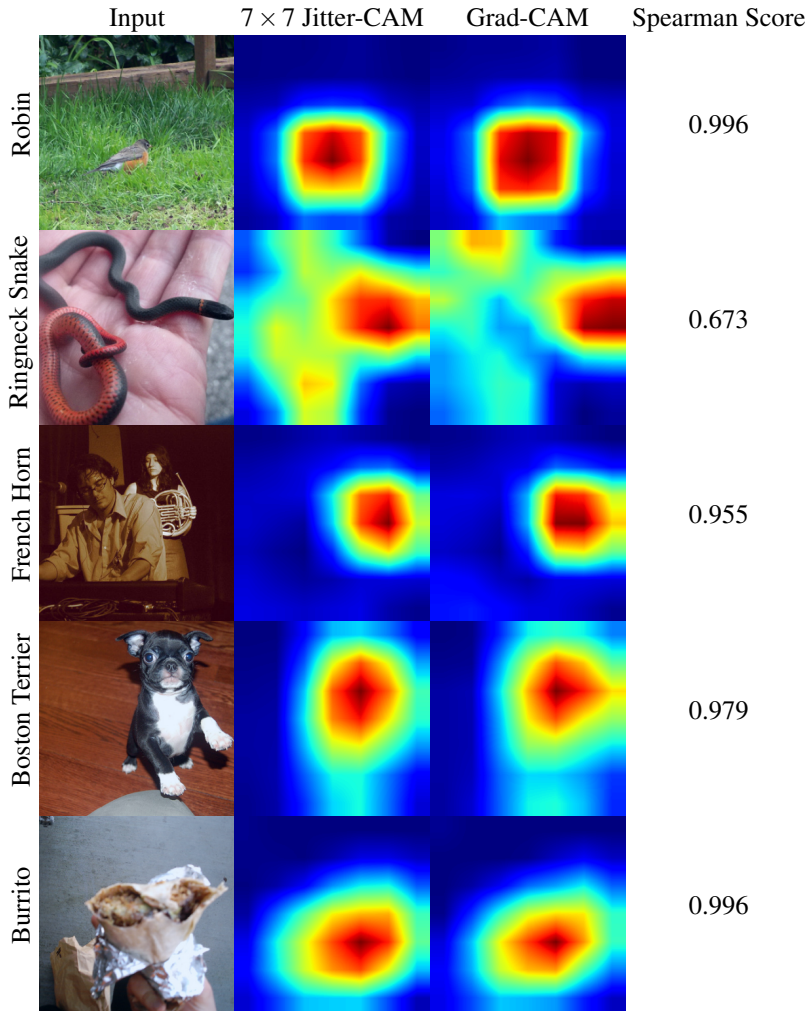


Figure 4: Comparison between Jitter-CAM explanations resized from  $5 \times 5$  to  $7 \times 7$  and their Grad-CAM counterparts. The Spearman score between the two explanations is also shown.

### 3 Faithfulness Baselines

In this section we show examples of the expanded explanations used in our faithfulness experiment. We show expanded Grad-CAM explanations in Figure 5, and our expanded random baseline in Figure 6. Note that for Grad-CAM any multiplication factor above  $\times 1$  (the original explanation) results in a better score for both the Average Drop (AD) and Increase In Confidence (IIC) metrics. For the random baseline, anything above  $\times 1.5$  performs better than Grad-CAM.

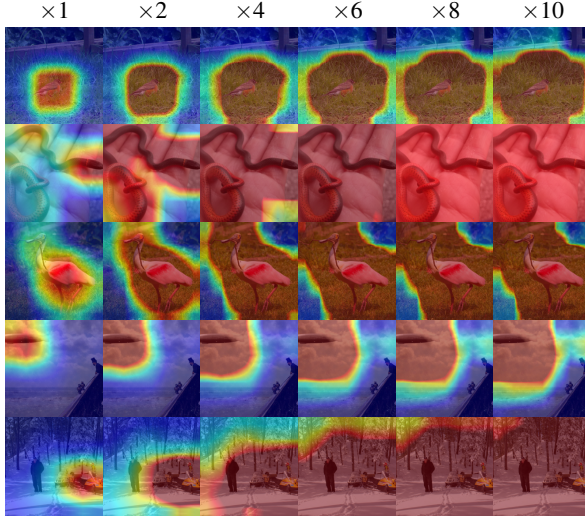


Figure 5: Example of expanded Grad-CAM explanations used in the faithfulness experiment.

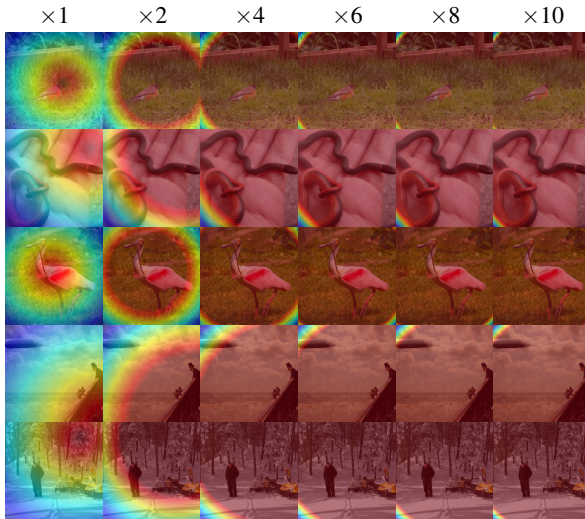


Figure 6: Example of expanded random baselines used in the faithfulness experiment.

## 4 Guided-CAM

As briefly discussed in the main paper, we found that Guided Jitter-CAM outperforms both Guided Grad-CAM and Guided Backpropagation when using ResNet50. In Figure 7 we show explanations generated using the three techniques. Results for the insertion and deletion metrics are shown in Table 1, while weak localisation results are shown in Table 2.

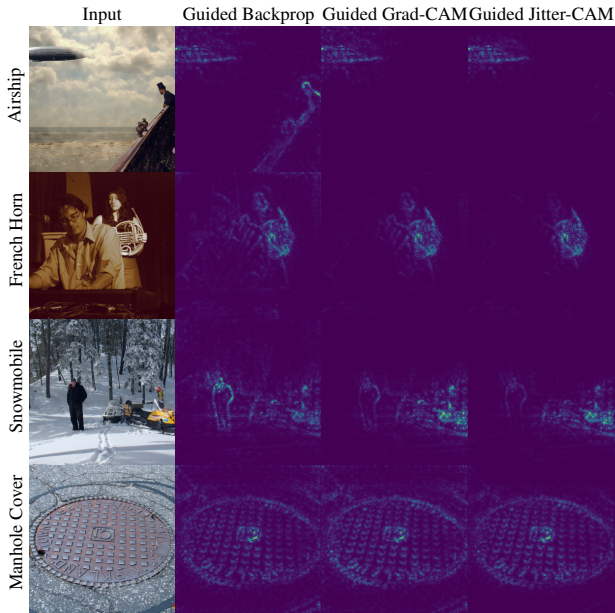


Figure 7: Examples of explanations created using guided methods.

Method	ResNet50	
	Del	Ins
Guided Backpropagation	0.074	0.416
Guided Grad-CAM	0.072	0.467
Guided Jitter-CAM	<b>0.071</b>	<b>0.489</b>

Table 1: AUC for local accuracy metric. Del: Lower is better. Ins: Higher is better.

Method	ResNet50		
	Val	Mea	Eng
Guided Backpropagation	56.49	51.63	52.49
Guided Grad-CAM	50.43	44.78	45.82
Guided Jitter-CAM	<b>47.42</b>	<b>41.86</b>	<b>43.09</b>

Table 2: Weak-localisation results as % of localisation error. Lower is better.



## 5 Example Patches

In this section we show the patches and explanations used to generate the Jitter-CAM explanation shown in Figure 8. We show image patches in Figure 9, and the corresponding  $7 \times 7$  Grad-CAM explanations in Figure 10.

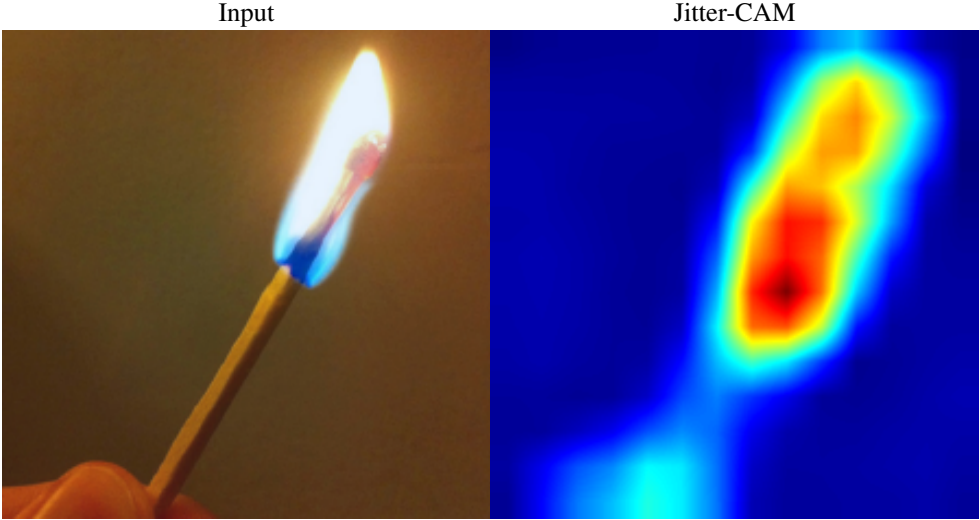


Figure 8: Input image alongside corresponding Jitter-CAM explanation that we show patches for in this section.

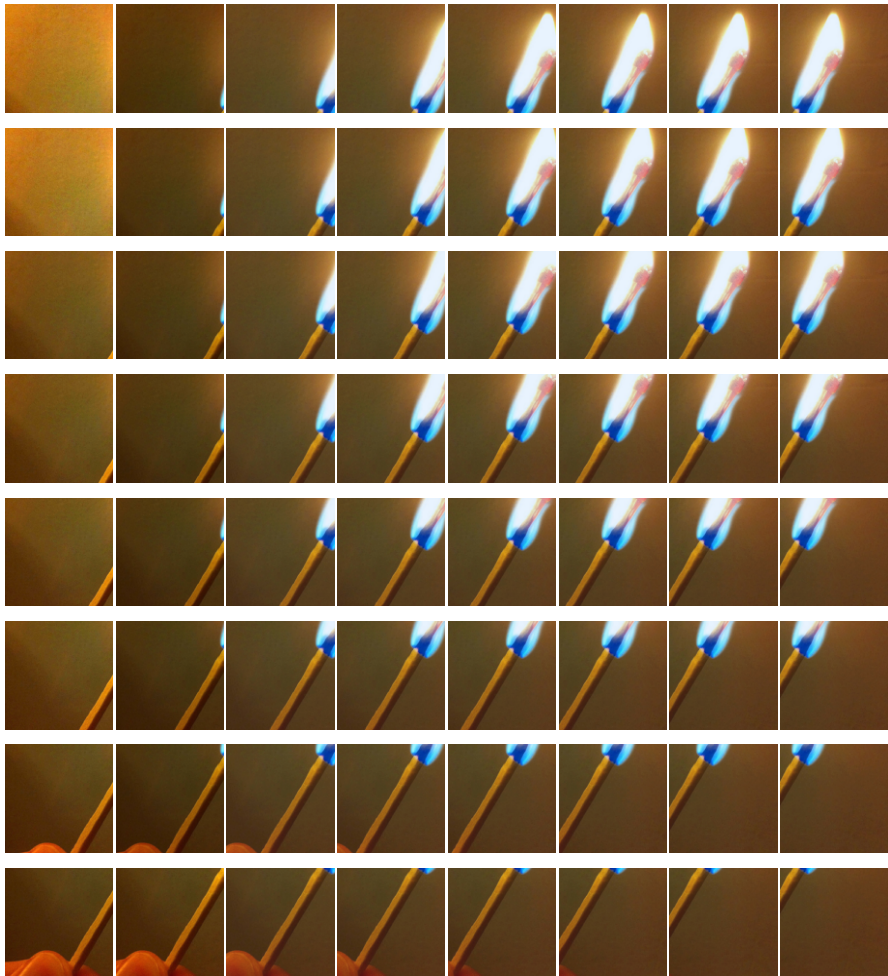


Figure 9: Example of the individual image patches used to make an explanation.

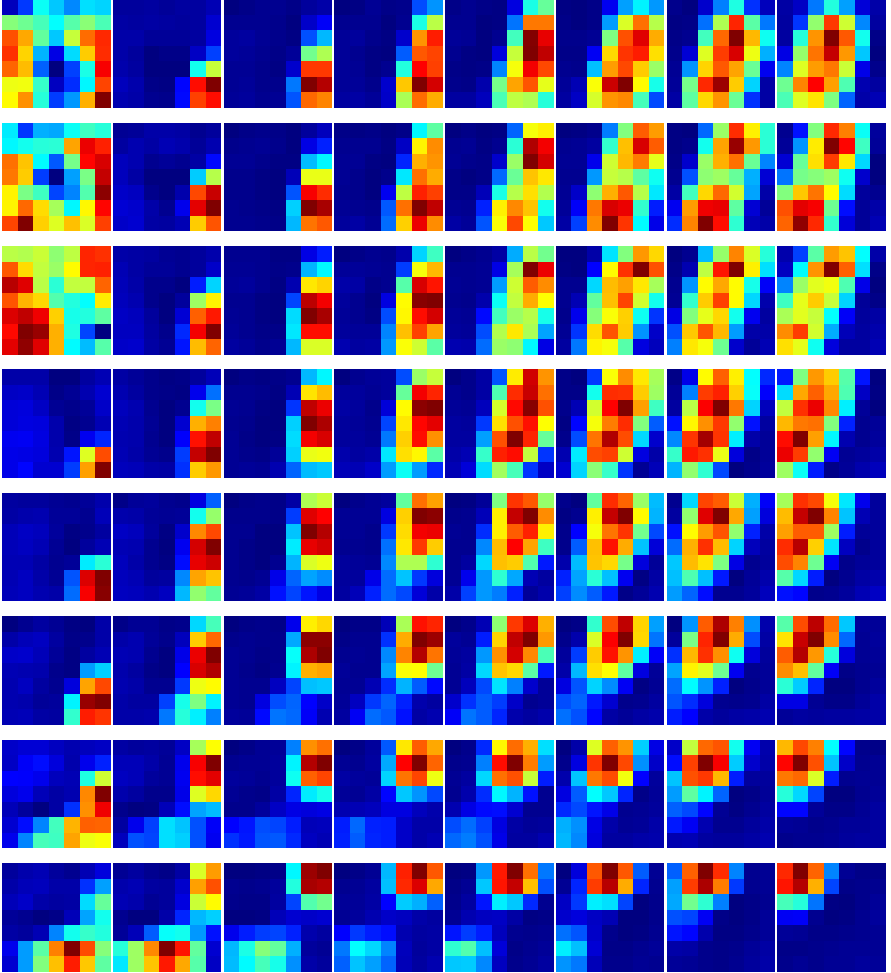


Figure 10: Examples of the  $7 \times 7$  Grad-CAM explanations created from the image patches. In each explanation, red is the highest scoring values in that patch, not the highest values across all patches.