

Grad-Cam

姓名：李紘宇

1 Grad-Cam 用途

以往的 Cam 必須重新訓練模型，才可以可視化模型關注的部分，而 Grad-Cam 運用反向傳播來計算特徵圖的重要性，不須重新訓練，即可得知模型關注的特徵。

2 實現步驟

假設網路分類器的輸出為 m 個神經元的全連接層， m 個神經元的輸出分別是 $z = [z^1, \dots, z^c, \dots, z^m]$ 。假設我們要獲取第 c 類的 Grad-Cam，由以下步驟可以得到：

1. 前向傳播得到該目標層第 k 通道的特徵圖 A^k 。
2. 運用反向傳播求特徵圖上 (i,j) 點上的導數 $\frac{\partial z^c}{\partial A^k_{i,j}}$ 。
3. 再做全局平均池化，得到 $a_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial z^c}{\partial A^k_{i,j}}$ ， $Z = H \times W$ ， H 和 W 為特徵圖的高和寬。
4. 最後得到 gradcam， $L^c_{gradcam} = RELU(\sum_k a_k^c A^k)$

3 原理

梯度 $\frac{\partial z^c}{\partial A^k_{i,j}}$ 指出了參數調整的方向，因此可以同時表明特徵圖的重要性，梯度值越大，代表此特徵的影響越大。因此當我們對一張特徵圖的每一點的梯度作平均，就可以得到這張圖的重要性 a_k^c ，最終我們可以將每張特徵圖 A^k 乘上它的重要性 a_k^c 後相加，得到 gradcam。

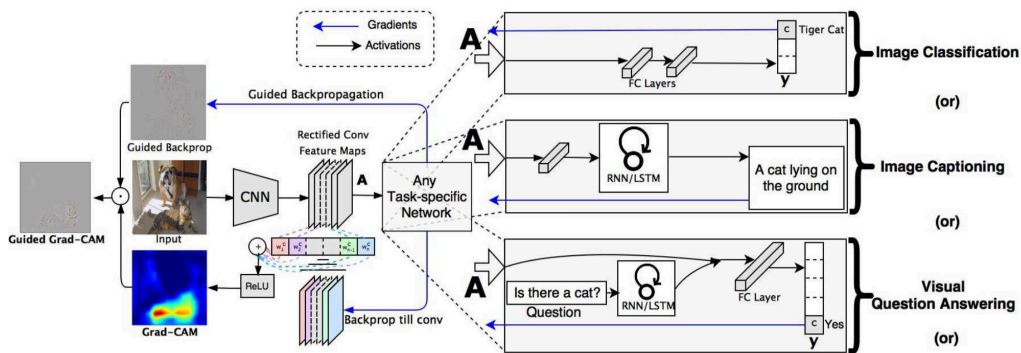


Figure 1: Grad-Cam 流程圖