WEEK 4. Visualization vs. Interpretability

	Summary	注意事项	Model	Activation / Gradients	comments
Lab 1	CAM - Fashion MNIST	example	example	Activation: Define model - Train - CAM model - CAM map	
Lab 2	CAM - Cats vs. Dogs	label没有one-hot encoding	Self-define sequential	-3 layer: Feature (3, 3, 128) —> (28, 28, 128) -1 layer: Weights (128, 10) —> (128,) CAM: Feature X Weights (28, 28)	
Lab 3 作业	Saliency - Cats vs. Dogs Saliency	TF和TF_hub有 compatibility问题 label有做one-hot	Inception V3 (from TF_Hub)	Calculate the Gradient w.r.t. input image **** a. tape.watch(inputs) b. Loss is calculate from expected_output + predictions	1. 什么是gradient 2. 要复习哪一课? - Course 2. WEEK 1. Differentiation
	- Cats vs. Dogs		sequential model	- categorical_crossentropyc. gradients = tape.gradient(loss, inputs)	and Gradients
Lab 4	GradCam - Cats vs. Dog	training慢,在colab 上快	VGG16 functional	Calculate the Gradient w.r.t. input image **** a. tape.watch(conv_output_values) b. Loss is calculate from actual_label + pred_prob (vs. smoothing) - binary cross entropy - loss = -1 * (actual_label * tf.math.log(pred_prob + smoothing) +	
				What it do? *****	
			CAM	Final layer of convolutions (features). Stack them on top of each others to get the overall map and then scale them up to superimpose it on the original image. So we could get an approximation of which part of the image ended up generating the feature maps.	
			Saliency	Where the pixels were found that were the most impactful to the final classification. And we did that by looking at the gradients of the final layers to see which ones had the steepest curve. And from there, we could figure out their location and plot them on the original image.	
			GradCAM (Gradient- weighted CAM)	Gradient-weighted Class Activation Map (GradCAM): It is a Class Activation Map (CAM) that uses the gradients of the final activations. Similar to saliency map, it uses gradients to find the pixels impacting the final classification	