This paper aims to quantitatively explain the rationales of each prediction that is made by a pre-trained convolutional neural network (CNN). We propose to learn