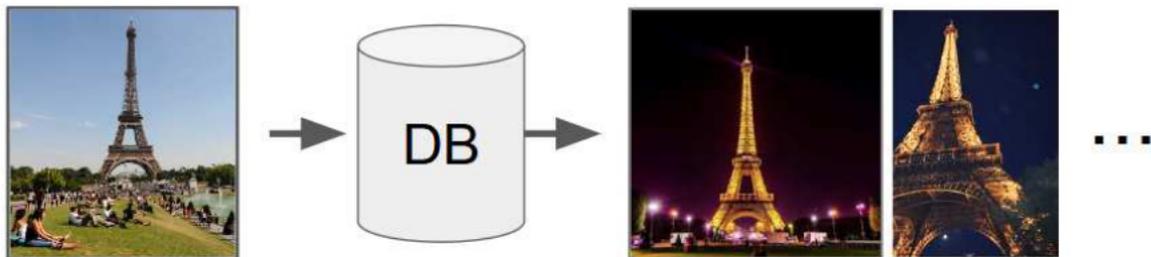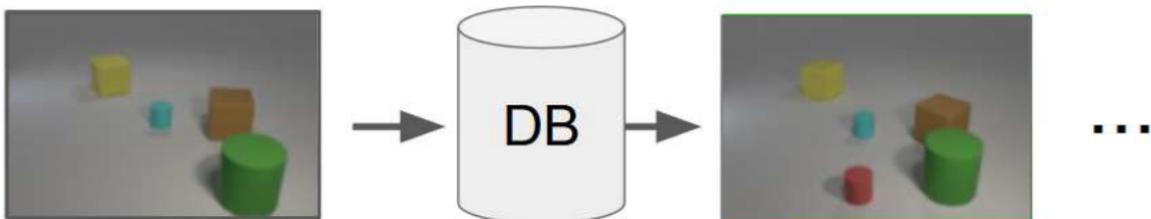**Image retrieval with text modifiers**

In this project, we will study the task of image retrieval, where the input query is specified in the form of an image plus some text that describes desired modifications to the input image. For example, we may present an image of the Eiffel tower, and ask the system to find images which are visually similar, but are modified in small ways, such as being taken at night time instead of during the day.

However, to simplify the problem we will make use of CLEVR dataset, synthetically created dataset consisting of simple objects. We will start building on top of the paper Composing Text and Image for Image Retrieval - An Empirical Odyssey.