1 UC San Diego
2 MIT-IBM Watson AI Lab
3 Massachusetts Institute of Technology

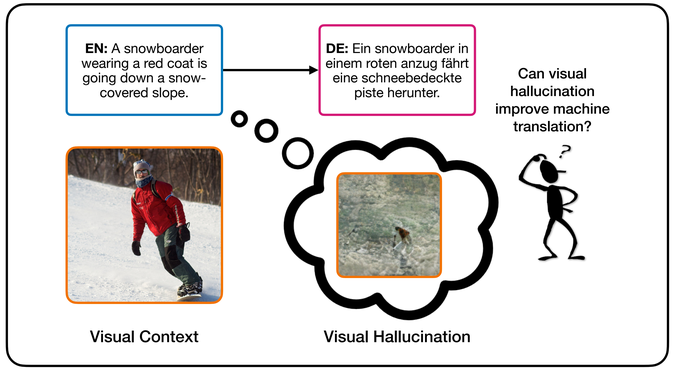
VALHALLA: Visual Hallucination for Machine Translation

Yi Li[1], Rameswar Panda[2], Yoon Kim[3], Chun-Fu (Richard) Chen[2], Rogerio Feris[2], David Cox[2], Nuno Vasconcelos[1]

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Overview

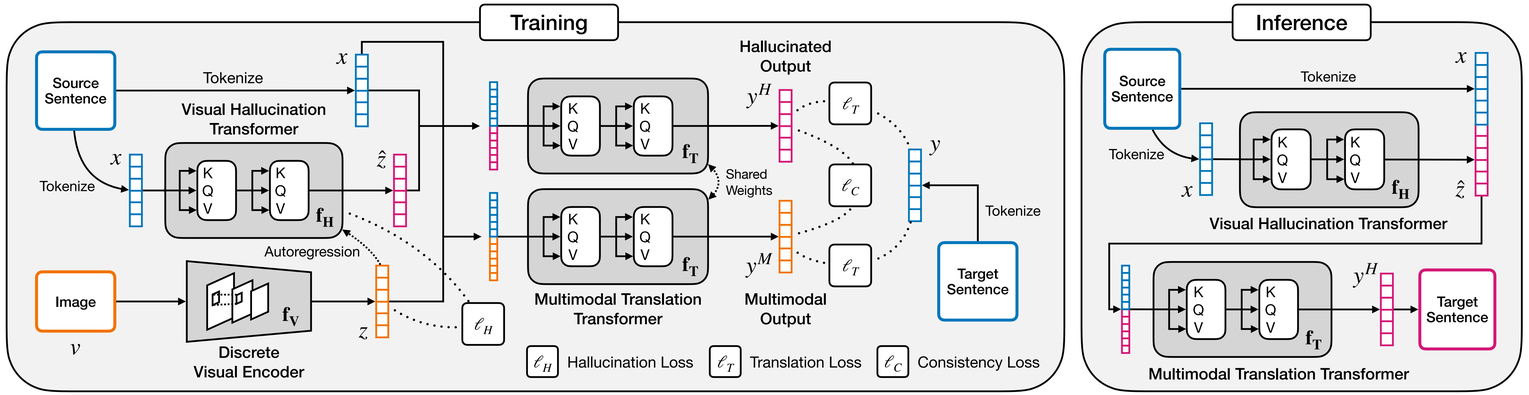
- Visual context has been exploited in **multimodal machine translation (MMT)** systems to improve translation performance
- In practice, inference with **paired image & text** not always realistic
- Can the ability to **hallucinate visual features** be used to improve language tasks, such as machine translation?
- We propose **VALHALLA**, a visual hallucination framework that learns from image-text correspondence during training, and predicts missing visual information for **text-only inference**



Code & Models

Project page

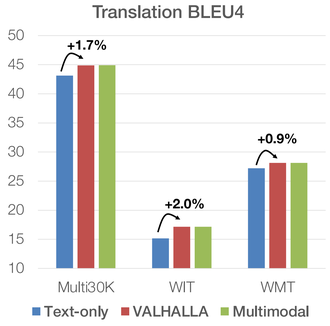
VALHALLA Architecture



- Stage 1 **Discrete visual encoder $\mathbf{f_V}$** tokenizes input image v into discrete sequence z
- Stage 2 **Visual hallucination $\mathbf{f_H}$** predicts visual tokens z from source sentence x
- Stage 3 **Multimodal translation $\mathbf{f_T}$** predicts target sentence y from concatenated input (x, z)
- **Text-only translation** by replacing z with hallucinated sequence $\hat{z} = \mathbf{f_H}(x)$

Experiments

- On multimodal datasets (**Multi30K, WIT**), VALHALLA improves performance over text-only transformer, matches MMT scores without requiring image input
- Training on text-only corpora (**WMT**) enabled through CLIP-based image retrieval
- Under **limited textual context**, VALHALLA recovers masked words, visual entities in source sentence
- Greater improvement for translating long sentences



Source EN	A boy wearing a red shirt digs into the sand with a yellow shovel.
Reference DE	Ein junge in einem roten shirt gräbt mit einer gelben schaufel im sand.
Text-Only	Ein junge, der ein rotes hemd trägt, wirft einen gelben ball in den sand. (*A boy wearing a red shirt throws a yellow ball in the sand.*)
VALHALLA	Ein junge in einem roten hemd gräbt mit einer gelben schaufel in den sand. (*A boy in a red shirt is digging in the sand with a yellow shovel.*)