**Motivation**
- Previous image captioning creates a general description of an image
- Users craft sentences based on their experiences using their own words

**Objective**
- Generate captions from image and user’s context

**Our Solution: CSMN**

**Context Sequence Memory Network**
- Multiple memory cell to condition different types of context information
- CNN memory I/O structure to jointly represent nearby ordered memory slots
- Sequence generation w/o RNN to capture long-term info without vanishing gradient

**InstaPIC-1M Dataset**
- Collected from Instagram
  - 1,124,815 unique posts and 6,315 unique users
- Goal:
  - Collect refined posts from Instagram
  - 27 general categories from Pinterest
  - 5 < caption length < 15, 50 < # posts per user < 1,000

**CSMN Architecture**
- Memory setup
- Prediction/update step
- Task1. Hashtag prediction
- Task2. Post generation

**Quantitative Results**
- Measured by both language and retrieval metrics

**User Studies via Amazon Mechanical Turk**
- General users’ preferences over the captions created by different methods for a query image

**Results**
- Hashtag prediction examples
- Post generation examples