

# **TERM-PAPER ASSIGNMENT**

**DUE: Tuesday, December 12, 2024, 4:00 PM**

**CSCI677: Advanced Computer Vision, Prof. Nevatia**

**Fall Semester, 2024**

This assignment is to write a term-paper on the topic of integrating vision and language modalities. It is intended to help you think about the techniques and capabilities of computer vision from a high-level perspective. Limit your report to two 8.5" x 11" (standard U.S. letter size) pages with 1" margins (top, bottom, left and right), 11 pt font and single line spacing. You may use up to a half-page, of the allowed two pages, for images and graphics if helpful. These images and graphics may be taken from published papers or websites as long as you provide the source citations.

In recent years, it has become possible to integrate vision and language modalities. You should attempt to answer the following questions.

- a) What motivates such integration and what benefits may be derived?
- b) What advances in processing of vision and language are enabling such integration?
- c) What are some successful integration examples that have already been reported?
- d) What are the limitations of the current approaches (if any)?

The above questions are broad. It is not intended that you will conduct extensive, external research to answer them. You can rely on the material presented in class though reading some of the referenced papers may help enhance your understanding. You may choose to focus on specific sub-topics, such as object detection or image captioning, where more progress may have been made or discuss broader topics in less detail.

Submit your term paper in pdf format in the assignments dropbox, as for the earlier assignments in this class.