

Ali Furkan Biten, R. Tito, Andres Mafla, Lluís Gomez, Marçal Rusiñol, C.V. Jawahar, et al. (2019). Scene Text Visual Question Answering. In *18th IEEE International Conference on Computer Vision* (pp. 4291–4301).