

Supplementary material

1. Additional qualitative results

We show additional qualitative results in this section.

Reconstruction Fig 1 and Fig 2 show reconstruction examples from four baselines in Crello and RICO datasets. Note the input document is from the test split, and the reconstruction might not be able to find the exact same image if the same image does not also appear in the training split (first row in Fig 1). While all baselines reconstruct to some extent, the oneshot Transformer model tends to reproduce the layout structure the best.

Interpolation Fig 3 and Fig 4 show the comparison of interpolation between two documents in Crello and RICO datasets. The comparison is between the autoregressive LSTM and the oneshot Transformer x4. The transition between two documents look smooth in both models, but the oneshot Transformer model seems to consistently pick a certain element (Fig 3).

Generation Fig 5 and Fig 6 show randomly generated examples from the oneshot Transformer x4 baseline.

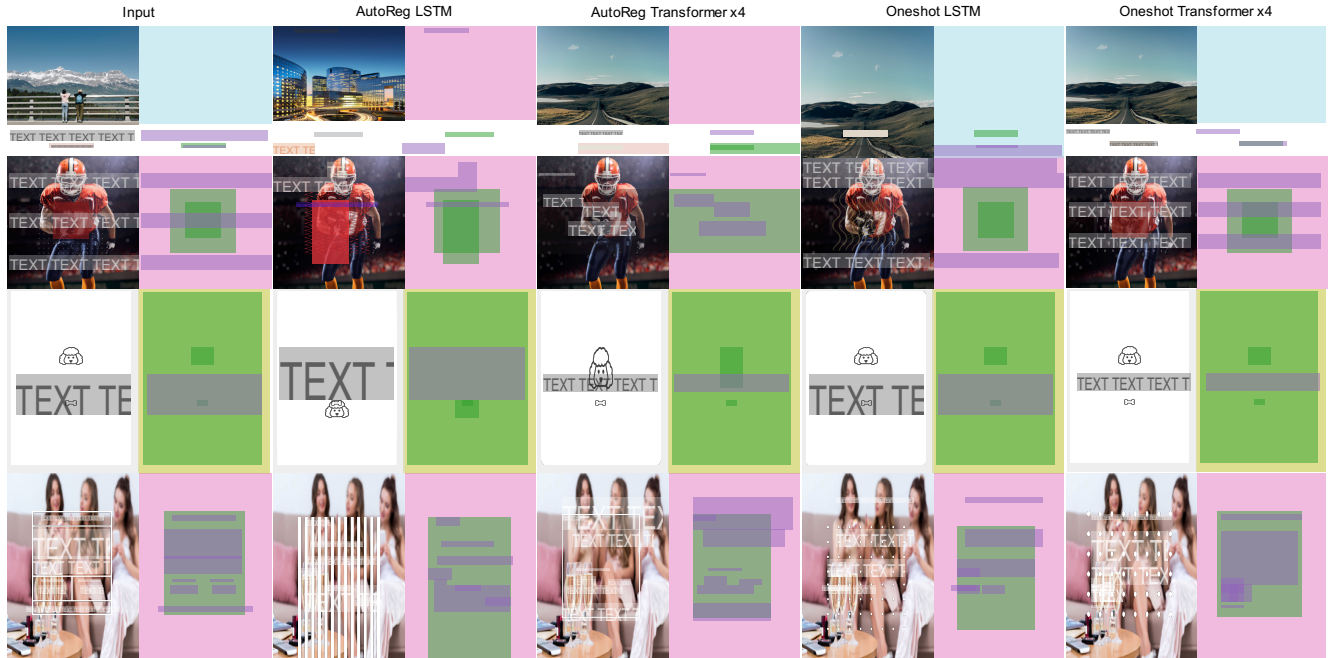


Figure 1. Additional reconstruction comparisons in Crello dataset. For each item, a left picture shows visualization of the design template with colored text placeholders and textured elements, and a right color map illustrates element types. The legend of types are the following: green = *vector shape*, magenta = *image*, purple = *text placeholder*, cyan = *masked image*, and yellow = *solid fill*.

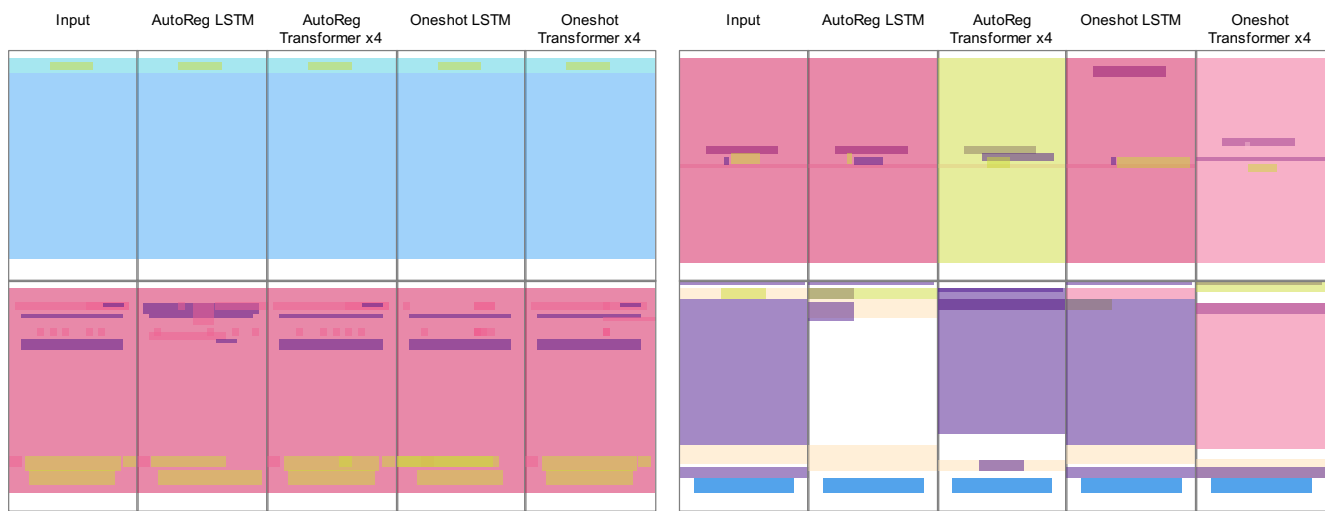


Figure 2. Additional reconstruction comparisons in RICO dataset.

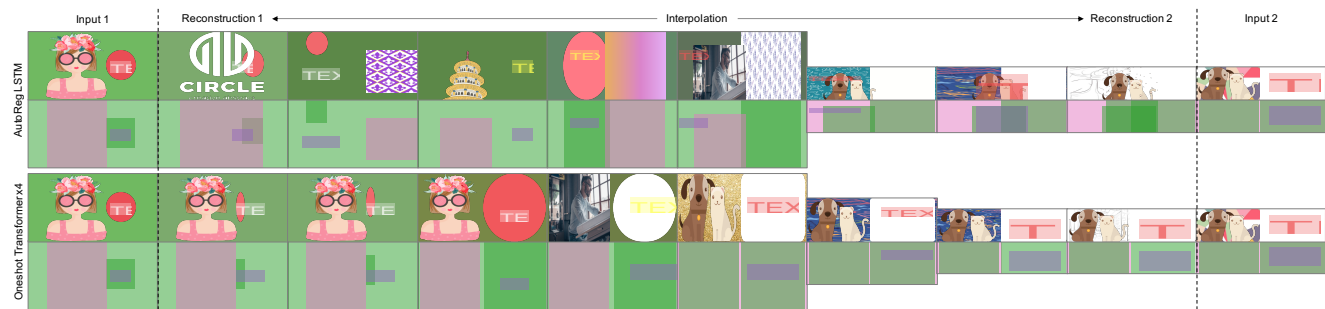


Figure 3. Additional interpolation examples from Crello dataset.



Figure 4. Interpolation examples from RICO dataset.

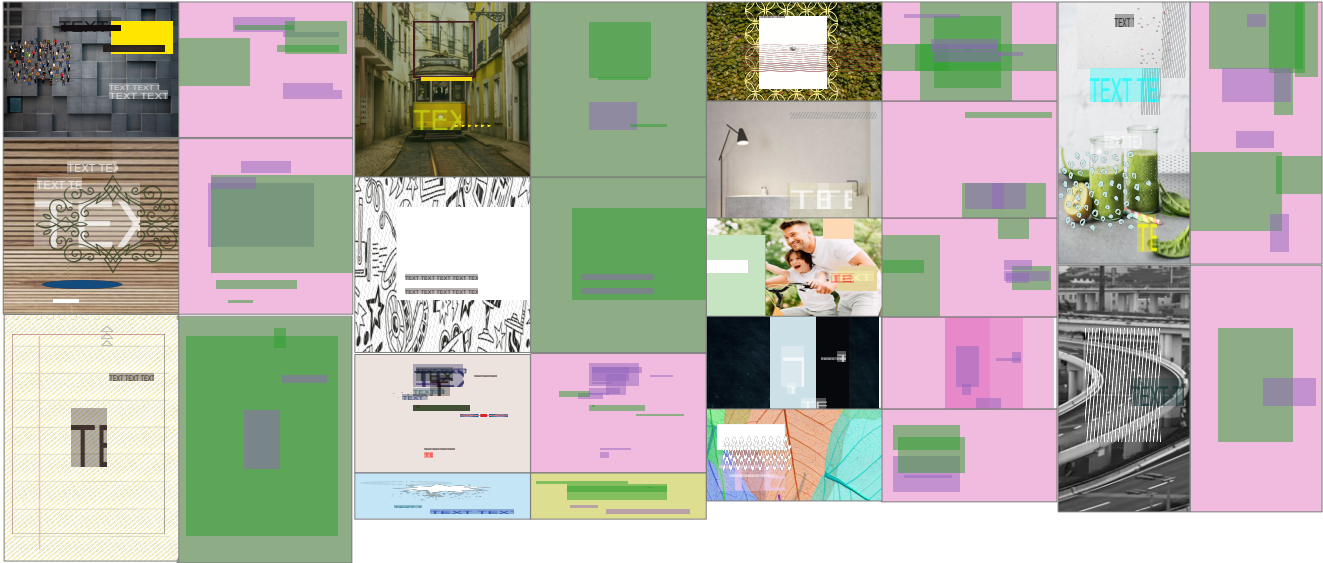


Figure 5. Random generation from Crello dataset.



Figure 6. Random generation from RICO dataset.