Appendix

A. Additional Qualitative Results

Figure 6. Additional qualitative examples of why attention matters. Similar to Fig. 3, we illustrate the $[CLS]$ token weights of the first attention layer. We show some examples for successful predictions (a, b, c) and some of the failure modes of our approach (d, e).