Nos modos convencionais de detecção de texto é utilizado um método de detecção que ocorre através de dois modelos separados para detecção e transcrição do texto em questão. Uma vez que alguns textos estão em imagens ou dispostos de formas não convencionais, é preciso detectar o posicionamento, início e fim do texto que será transcrito.

Nesse artigo, é proposto um novo modelo de transcrição de texto que utiliza a natureza multitarefas das arquiteturas de Redes Neurais Profundas, *Deep Neural Architectures.*

Foi construído um modelo capaz de detectar texto tanto de imagens quanto de um parágrafo de documento. Utilizando uma arquitetura baseada em 4 redes neurais de aprendizado profundo interligadas, cada uma realizando uma parte das tarefas de identificar, classificar as dimensões e transcrever o texto.

Através desse método foram obtidos diversos avanços na detecção e transcrição de texto convencionais, além de limitações no método descrito. Com sucesso, foi possível reduzir a propagação de erros que ocorria nos modelos convencionais, onde erros na etapa de detecção do texto acabam impactando a etapa de transcrição.