1.

a) The character embeddings don't need to be as large since a character can only represent so many things and meanings of words are attained when many characters are concatenated. Also it reduces computational complexity compared to higher dimensional characters, which is a good thing in character level models

b) Figure 2: character-based

$$\underbrace{vocab_{char} * e_{char}}_{\text{char embed}} + \underbrace{k * e_{char}}_{\text{conv}} \underbrace{+2 * (e_{word} * e_{word} + e_{word})}_{\text{projection}}$$

Figure 1:

$$vocab_{word} * e_{word}$$

c) The CNN approach allows for more efficient parallelization vs RNN. Also CNN can specify the number of filter for allowing the detection of many different features/aspects of the words

d) Max pooling will retain the strongest signal in the data, but discards other parts of data which may not be desirable.
Average pooling will retain all info, but the signal might get diluted if the matrices are sparse and the result may end up being very small.