

# Jawaban *Text Recognition*

## Soal

- Jelaskan arsitektur yang Anda buat untuk melakukan ekstraksi teks pada gambar!
- Apa yang kamu ketahui tentang CTC?
- Apa yang kamu ketahui tentang *attention*?

## Jawaban

### Jawaban a

(Arsitektur diperoleh dari [https://keras.io/examples/vision/captcha\\_ocr/](https://keras.io/examples/vision/captcha_ocr/), sehingga sebenarnya bukan saya yang buat. 😊) Secara garis besar, arsitekturnya menggunakan CNN (*convolutional neural network*) dan RNN (*recurrent neural network*) sehingga dapat disebut juga CRNN (*convolutional recurrent neural network*). Dari hasil pemrosesan CRNN tersebut didapatkan beberapa karakter yang berulang sehingga perlu diproses dengan CTC (*connectionist temporal classification*).

Untuk penjelasan lebih detail, CNN terdiri dari dua pasang *convolution layer* dan *max-pooling* yang digunakan untuk mengurangi dimensi dari input. Kemudian, dilanjutkan dengan RNN yang terdiri dari dua LSTM. Di sinilah mesin seakan-akan membaca tulisan tersebut dari kiri-ke-kanan dan menentukan huruf yang dimaksud. Outputnya dimasukkan ke *full-connected-layer (dense)* dan katanya diprediksi oleh CTC untuk mengurangi karakter yang berulang.

### Jawaban b

CTC adalah jenis *neural network* yang bertugas menangani masalah observasi berulang pada garis waktu. Intinya, dia mencoba mencari tahu nilai yang ditangkap oleh observasi yang berulang, misalkan membaca karakter yang bisa saja mendapatkan huruf berurutan yang sama pada beberapa wilayah gambar yang berurutan.

### Jawaban c

Secara bahasa, *attention* adalah perhatian. Ini adalah salah satu pendekatan dalam membaca tulisan selain CRNN, yaitu dengan mencari objek—diharapkan itu huruf—yang kelihatannya menarik dan mencoba mengidentifikasi karakter yang sesuai dengan objek tersebut. Hal itu dilakukan terus menerus dalam garis waktu (dalam OCR, dari kiri ke kanan).