* Trong compile(tối ưu) mình nên sử dụng cái loss hay optimizer nào ?
* Ai là bài toán cần tinh chỉnh, mỗi model loại data thì nó có 1 sự khác nhau nhất định.
* Cnn thì hay dùng cho ảnh, Rnn dùng cho câu chữ
* Optimizer thì khi data nhỏ có thể dùng gradient\_descent load 1 lúc tất cả cái data lên. Khi data lớn thì phải dùng stochastic gradient descent để tránh tràn ô nhớ nhưng cái learning\_rate của nó thì lại fix cứng nên nếu muốn thay đổi learning\_rate thì dùng adam, RMSProp
* Khi nào dùng padding=”same” ?
* Nó sẻ giúp lưu giữ lại data ở phía viền, nghĩa là bức ảnh nó lớn hơn. Trong trường hợp nhiều data quan trọng ở viền thì chúng ta dùng nó
* Số lượng maxsize, filter kernel size ?
* Best practice là số lượng datapoint nó phải tương ứng với số lượng paramaters
* Ví dụ kernel size thử 3 ,5
* Khi làm mạng chúng ta có thiết kế mạng theo chiều rộng, dài
* Gradient vanishing là gì ?
* Là đạo hàm của nó sẻ bằng 0 mà chúng ta sẻ ko học được gì
* Có nhiều vấn đề liên quan sigmoid có thể gây ra. Đạo hàm max của sigmoid là 0.25
* Train ML trên CPU sẻ ko nhanh bằng trên GPU được tuy nhiên củng có các cái khác tốt hơn TPU chuyên dụng cho AI
* Inference chỉ để chạy không train được
* Train trên TPU thì cần học sâu hơn ví dụ như datatype float 32, 64