

0407 연구노트

AI 비교 지표

AI Score

[Comparison and Benchmarking of AI Models and Frameworks on Mobile Devices](#)의 VII. Technical Description - D. AI Score 출처

- Valid Images Per Second

$$VIPS = \sum_{i=1}^n accuracy_i \times \frac{1}{time_i}$$

- Valid FLOPs Per Second

$$VOPS = \sum_{i=1}^n accuracy_i \times FLOPs_i \times \frac{1}{time_i}$$

$accuracy_i$ 는 i 번째 validation accuracy, $time_i$ 는 i 번째 사진에 대한 추론 시간 평균, $FLOPs_i$ 는 i 번째 모델의 FLOPs

FLOPs

FLoting point **OP**erations의 약자로, 부동소수점 연산량을 나타내는 지표
사칙연산을 포함하여 $\sqrt{\quad}$, \log , \exp 등의 연산을 각각 1회 계산

****FLOPS****는 **FL**loating **OP**erations **P**er **S**econd로 초당 얼마나 많은 연산을 처리할 수 있는지를 나타내는 하드웨어 지표로, 구분해야함

FLOPs 계산

- Fully Connected Layer

$inputlayer = n$, $outputlayer = m$ 일 때, m 개에 대한 내적이 n 번 발생
 $FLOPs = (2n - 1) \times m$

- Convolutional Layer

$H \times W \times C$ 의 입력 이미지에 $K \times K$ 커널을 적용한다고 가정

$$FLOPs = H \times W \times C_{in} \times C_{out} \times K \times K$$

연산량이 상당하기 때문에 Depthwise Seperable Convolution 등등의 경량화 기법이 등장