TANGO

트랜스포머 NAS

성명 신익수 소속 한국전자통신연구원

SUBJECT

인공지능 기술의 대중화 (AI Democratization)를 위한
TANGO 커뮤니티 3회 컨퍼런스

주관 ETRI 한국전자통신연구원 (TANGO) 주최 과학기술정보통신부 IITP 정보통신기획평가원

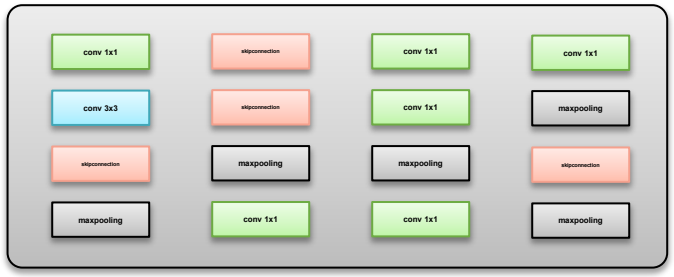
후원 Line lablup weda tesla System (사)한국인공지능협회 SNUH 서울대학교병원 고려대학교 KOREA UNIVERSITY 경희대학교 KYUNG HEE UNIVERSITY 중앙대학교

목 차

NAS 기술 발전 History

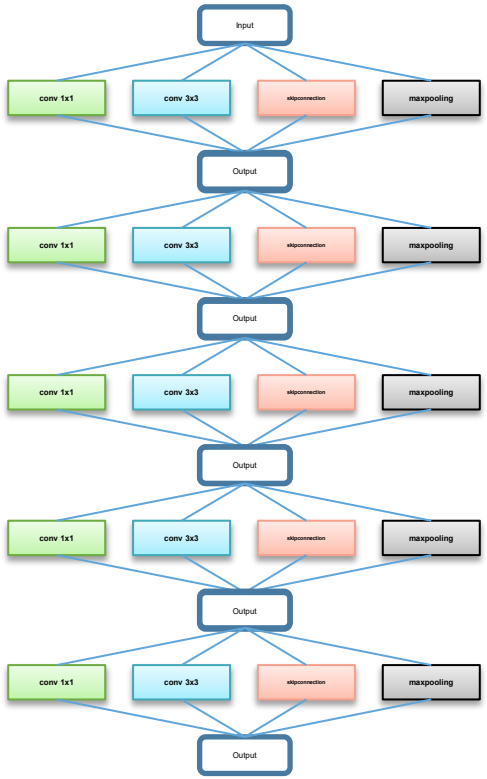
Classic NAS

Generation *N*



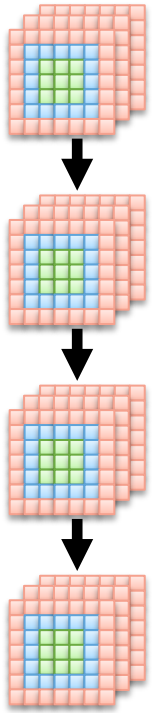
Generation *N+1*

OneShot NAS



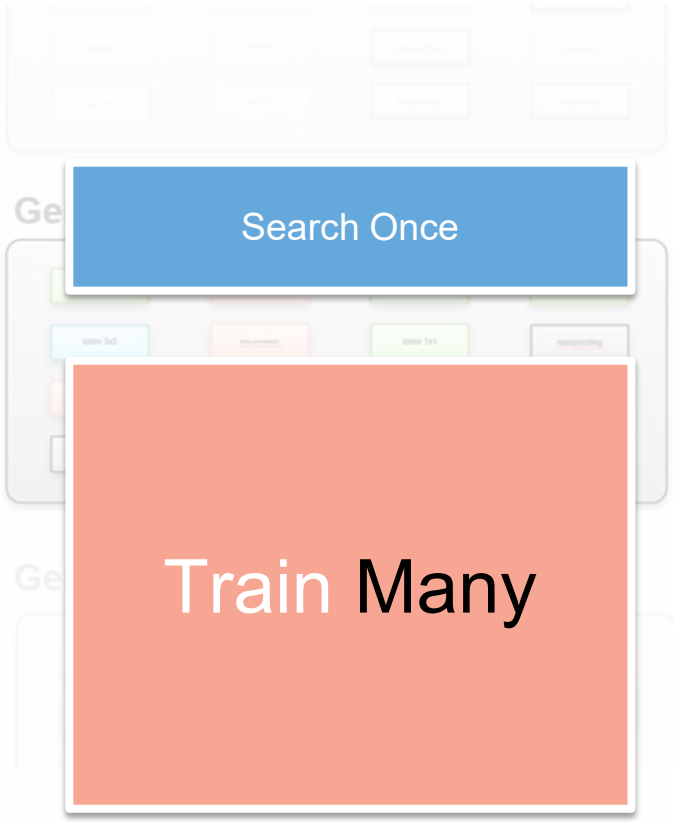
Once-for-all Style

OneShot NAS

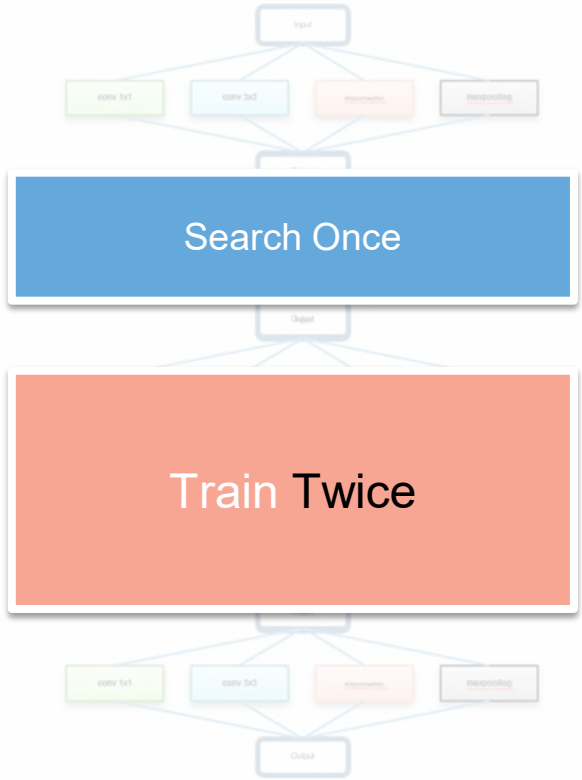


NAS 기술 발전 History

Classic NAS

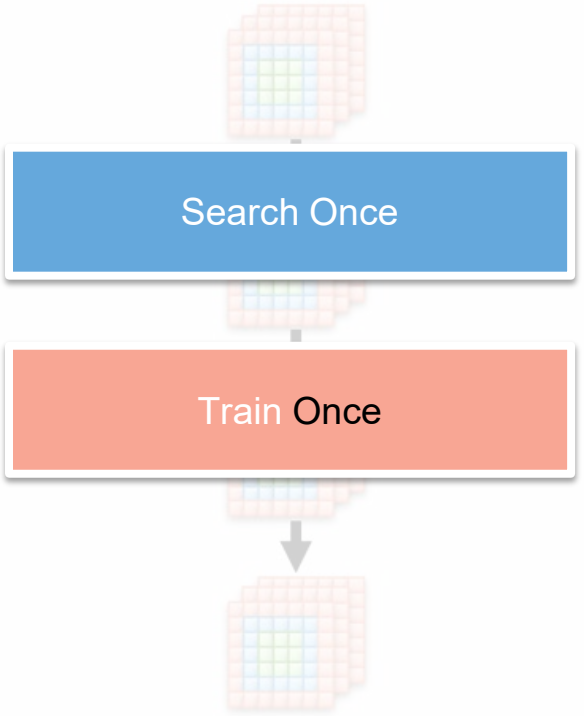


OneShot NAS



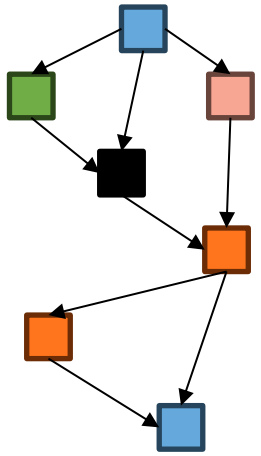
Once-for-all Style

OneShot NAS



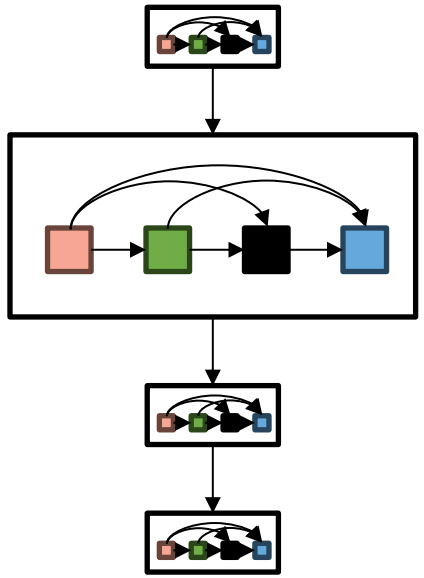
NAS 기술 발전 History

Classic NAS



다양한 구조 생성 가능

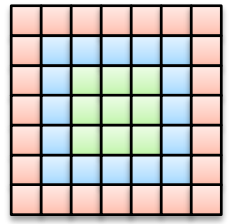
OneShot NAS



블록 단위 구조

Once-for-all Style

OneShot NAS



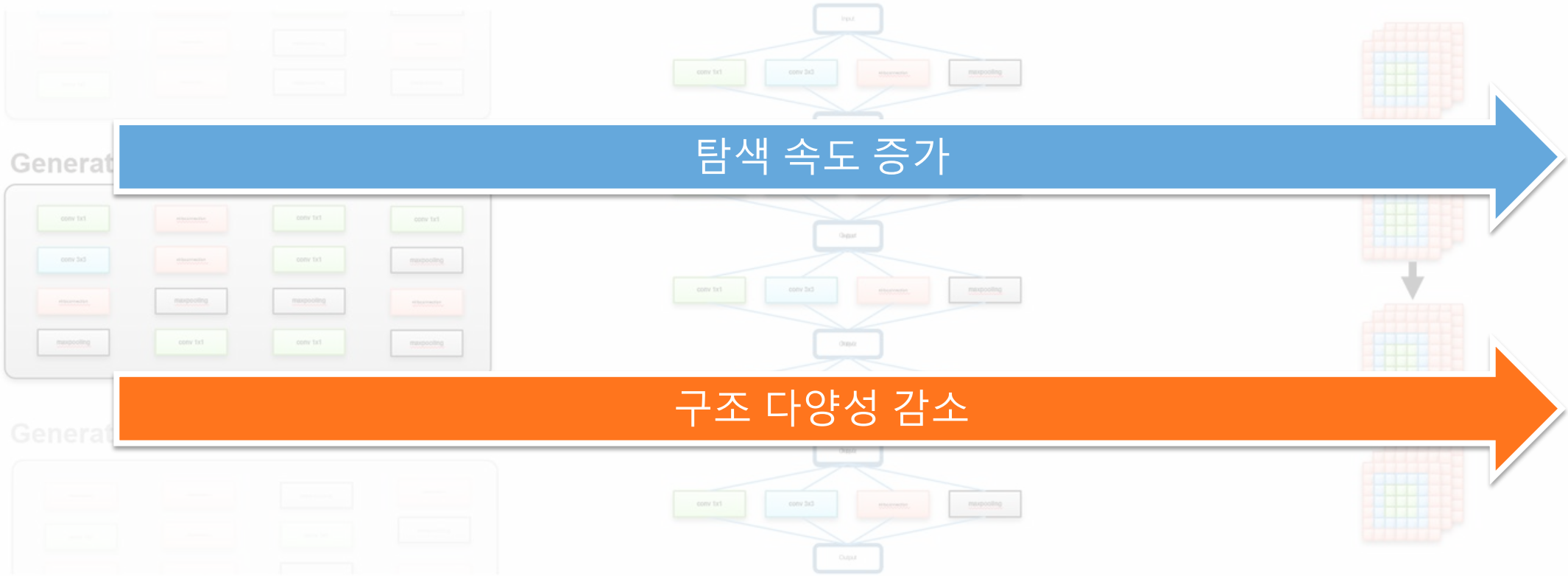
연산의 크기만 조절

NAS 기술 발전 History

Classic NAS

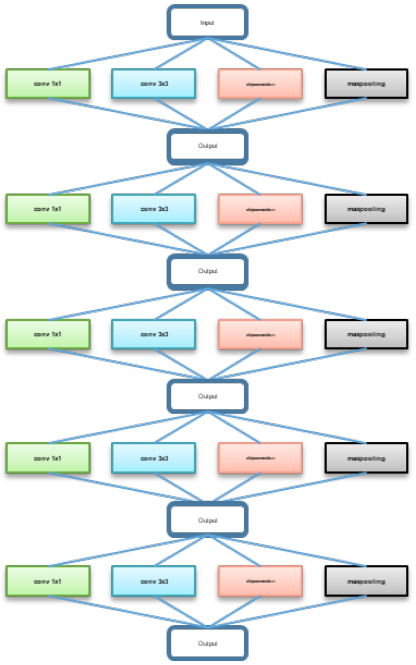
OneShot NAS

Once-for-all Style
OneShot NAS

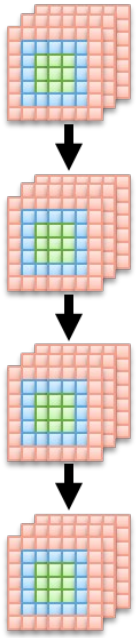


연구 문제 설정

OneShot NAS



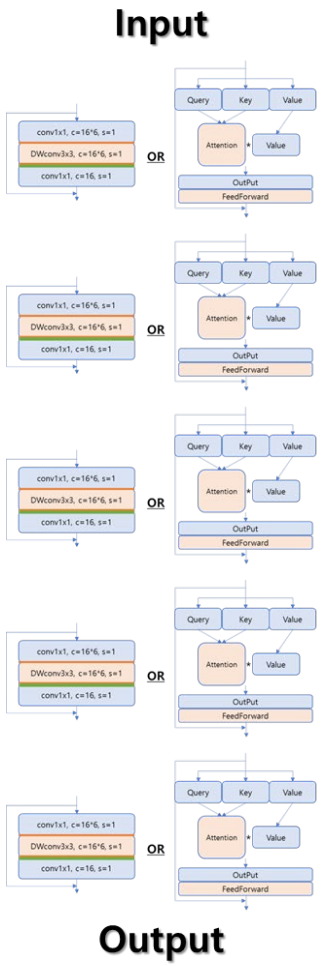
Once-for-all Style
OneShot NAS



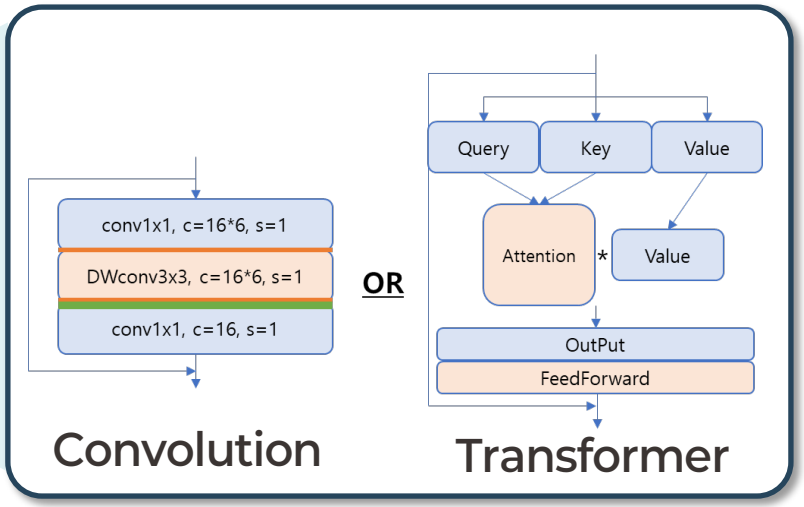
구조 다양성

고속 탐색

연구 문제 설정



각 레이어마다 연산자 선택권



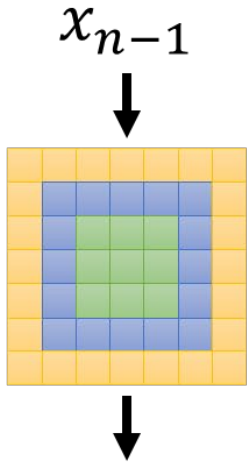
연구 문제 설정

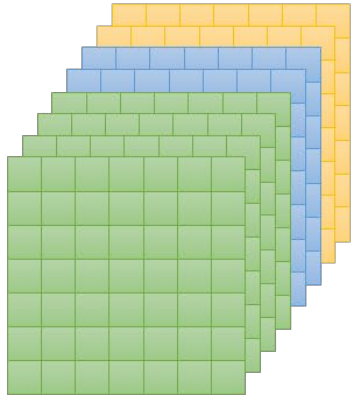
고속 탐색의 비결은 연산자들 간의 가중치 공유 구조

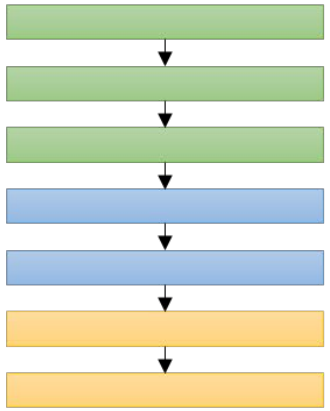
: core part

: extra part 1

: extra part 2







$$O_{small} \subset O_{bigger}$$

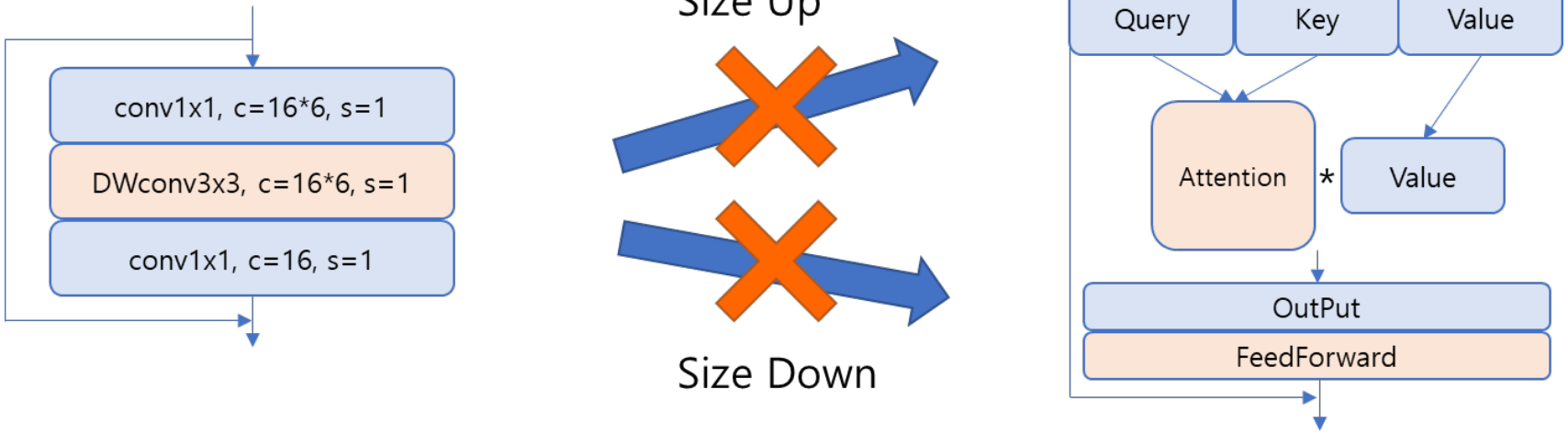
Filter Size

The number of channels

The number of layers

연구 문제 설정

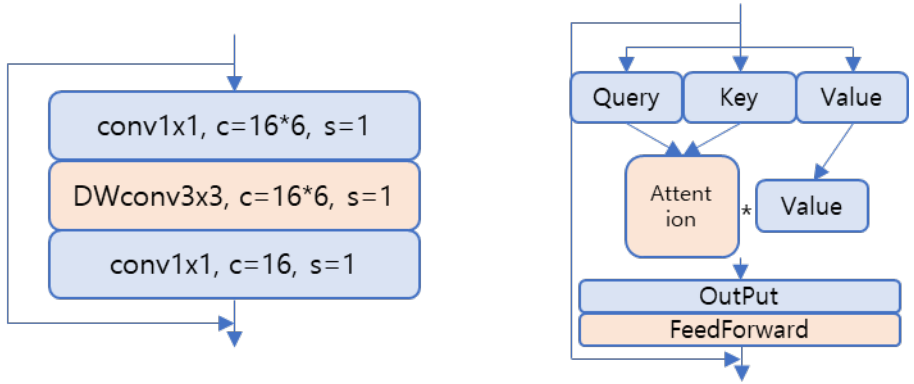
다른 종류의 연산자들 간에는 가중치 공유 구조 불가



연구 문제 설정

Naïve Approach

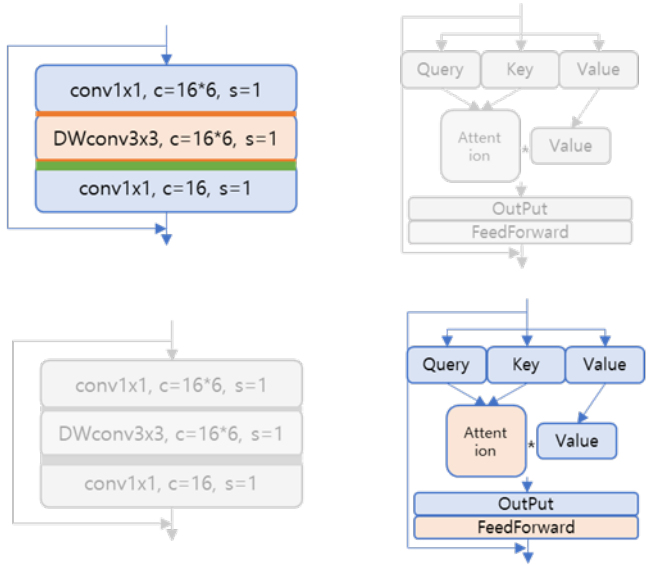
Approach 1. Convex Combination



$$\alpha * y_{conv}(x) + (1-\alpha) * y_{att}(x)$$

$$where\ \alpha \in (0,1), trainable$$

Approach 2. Random Sampling

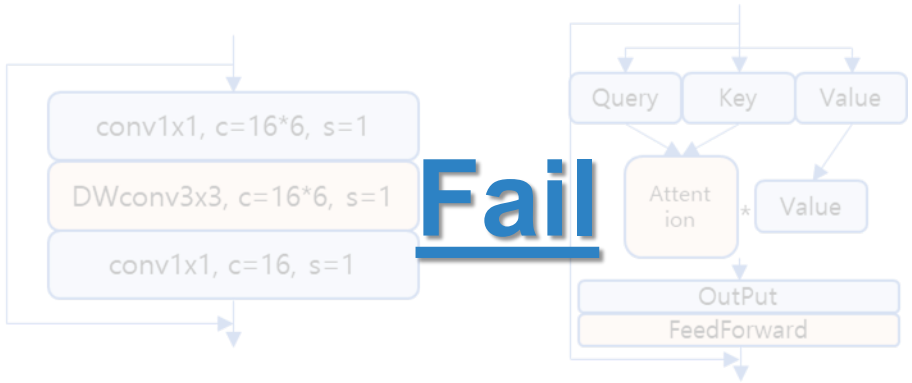


$$\alpha \sim UniformDistribution$$

연구 문제 설정

Naïve Approach

Approach 1. Convex Combination



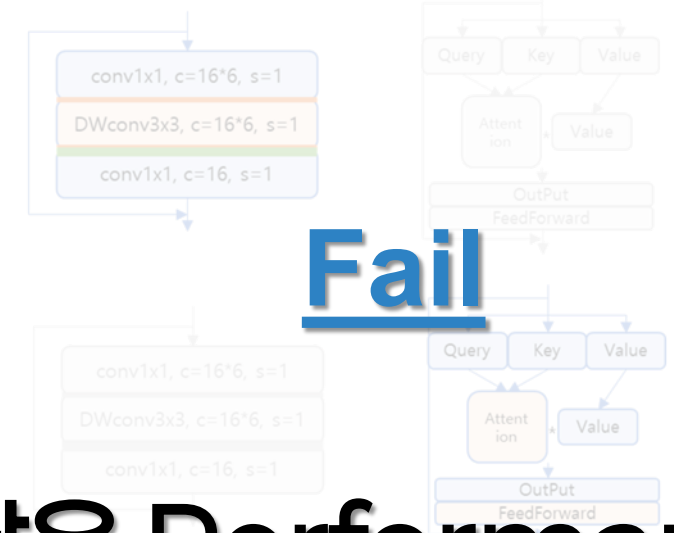
Fail

$\alpha * y_{n-1} + (1-\alpha) * y_n$

$where\ \alpha \in (0,1), trainable$

α의 수렴 실패

Approach 2. Random Sampling



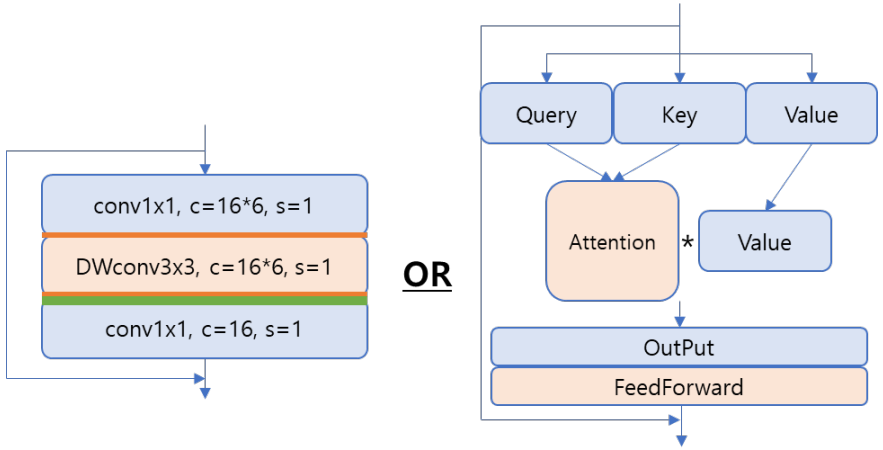
Fail

$\alpha \sim UniformDistribution$

낮은 Performance

신규 NAS 기술

매 스텝마다 연산자 가중 샘플링 (Weighted Samling)



$F(x_{n-1})$ $G(x_{n-1})$

샘플링 확률 실시간 업데이트

$$x_n = \begin{cases} (\alpha + T) \cdot \mathrm{F}(x_{n-1}) & \text{if } s=1 \\ T: 1 - \alpha.detach() & \\ (\beta + T) \cdot \mathrm{G}(x_{n-1}) & \text{if } s=0 \\ T: 1 - \beta.detach() & \end{cases}$$

$$s \sim \mathrm{Bernoulli}(\alpha)$$

$$\alpha + \beta = 1, \quad \alpha, \beta \in (0,1)$$

α: Convolution 샘플링 확률
β: Transformer 샘플링 확률

신규 NAS 기술

$$\boldsymbol{\alpha \cdot F(x_{n-1}) + (1-\alpha) \cdot G(x_{n-1})}$$

$$\boldsymbol{(\alpha + T) \cdot F(x_{n-1})}$$

$$\mathrm{T}:\ 1 - \alpha.\mathrm{detach}()$$

forward:

$$x_n = (\alpha + (1-\alpha)) * \mathrm{ops}(x_{n-1})$$

$$= \mathrm{ops}(x_{n-1})$$

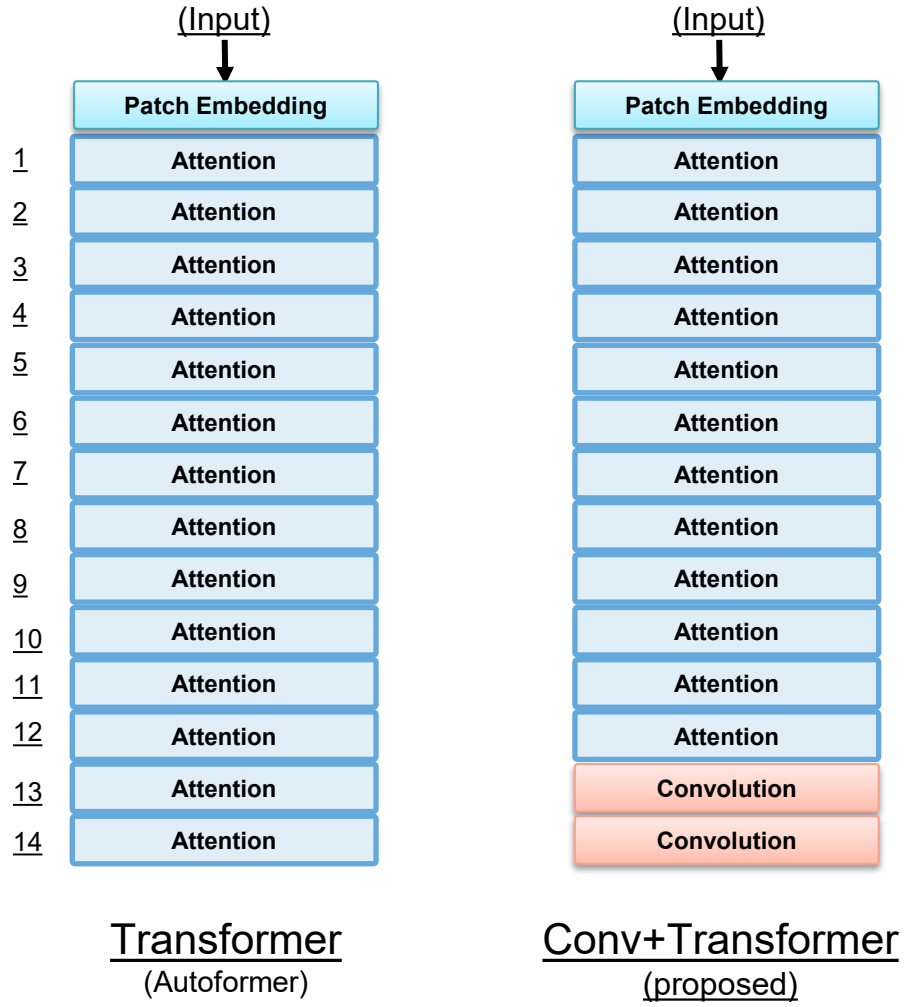
backward:

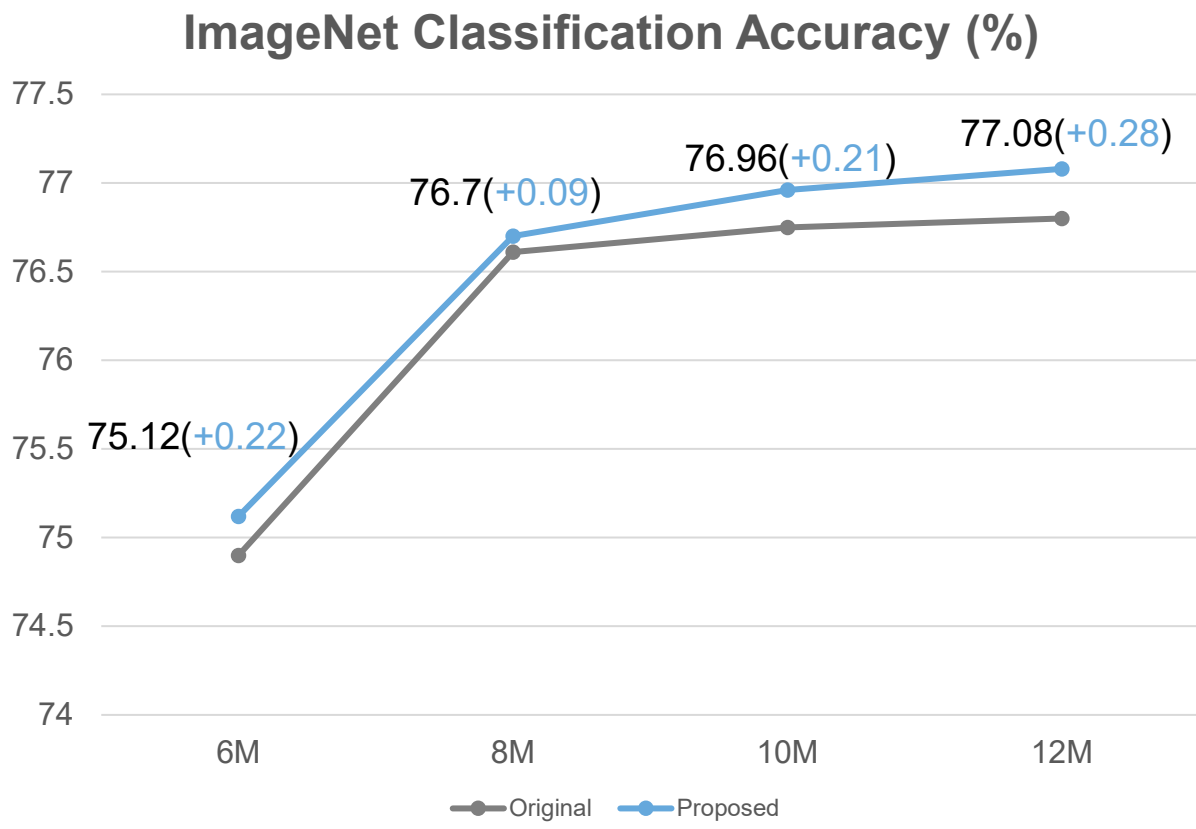
$$\frac{\partial x_n}{\partial \alpha} = \mathrm{ops}(x_{n-1}) \ + \ 0 - 0$$

$$= \mathrm{ops}(x_{n-1})$$

신규 NAS 기술

실험 결과





TANGO

TANGO

감사합니다.

주관 ETRI 한국전자통신연구원 (TANGO) 주최 과학기술정보통신부 IITP 정보통신기획평가원

후원 Line labiup weda tesla system (사)한국인공지능협회 SNUH 서울대학교병원 고려대학교 KOREA UNIVERSITY 홍익대학교 HONGIK UNIVERSITY CAU 중앙대학교