2024 겨울 세미나

2024.02.16

Sogang University
Vision & Display Systems Lab, Dept. of Electronic Engineering

Presented By
김동규

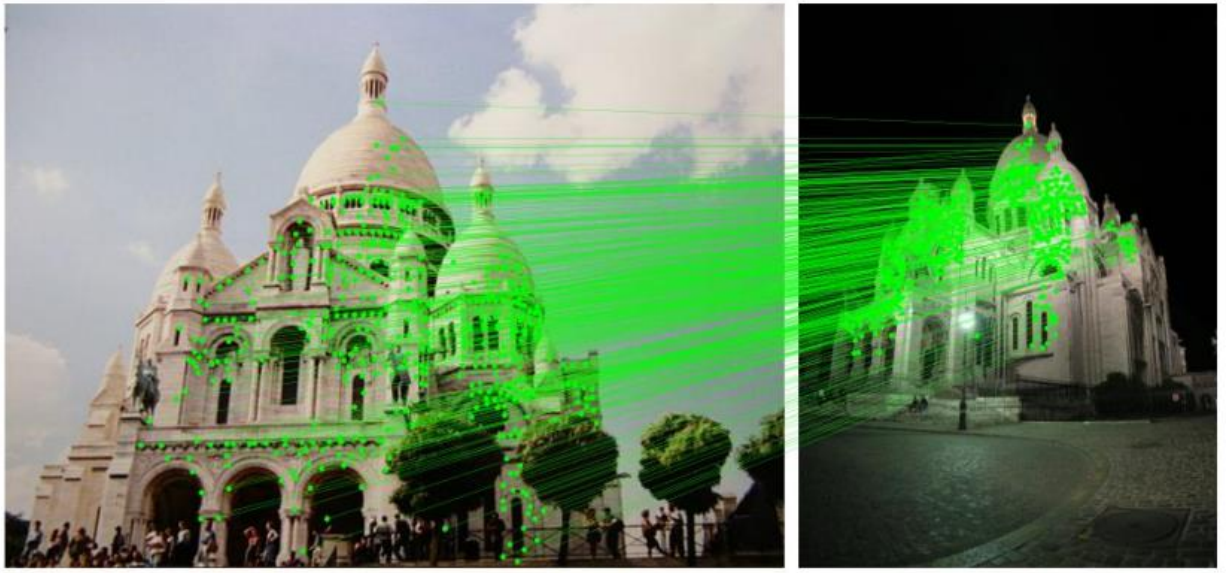
Outline

- Background
  - What is local feature matching
  - SuperGlue: Learning Feature Matching with Graph Neural Networks (CVPR 2020)

- LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

Background

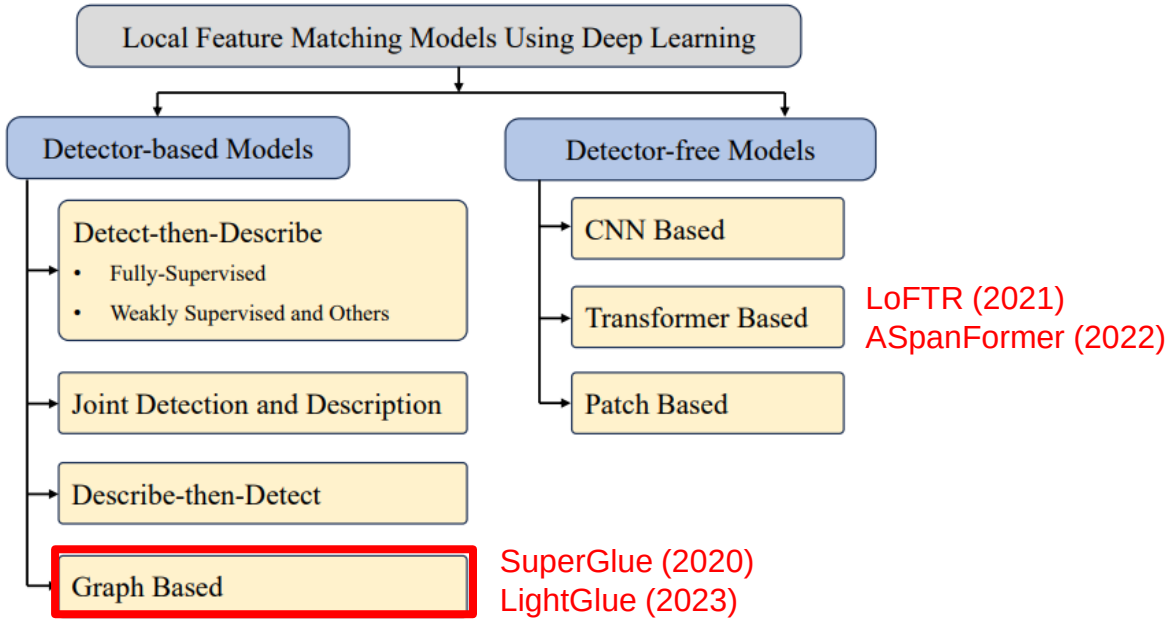
- What is local feature matching
  - 서로 다른 image 에서 correspondences 를 찾는 과정
  - Image fusion, visual localization, SfM (Structure from Motion), SLAM (Simultaneous Localization and Mapping) 등 많은 분야에서 이용
  - (Challenge) matching accuracy와 robustness 상승



< 두 image 사이의 feature matching 예시 >

Background

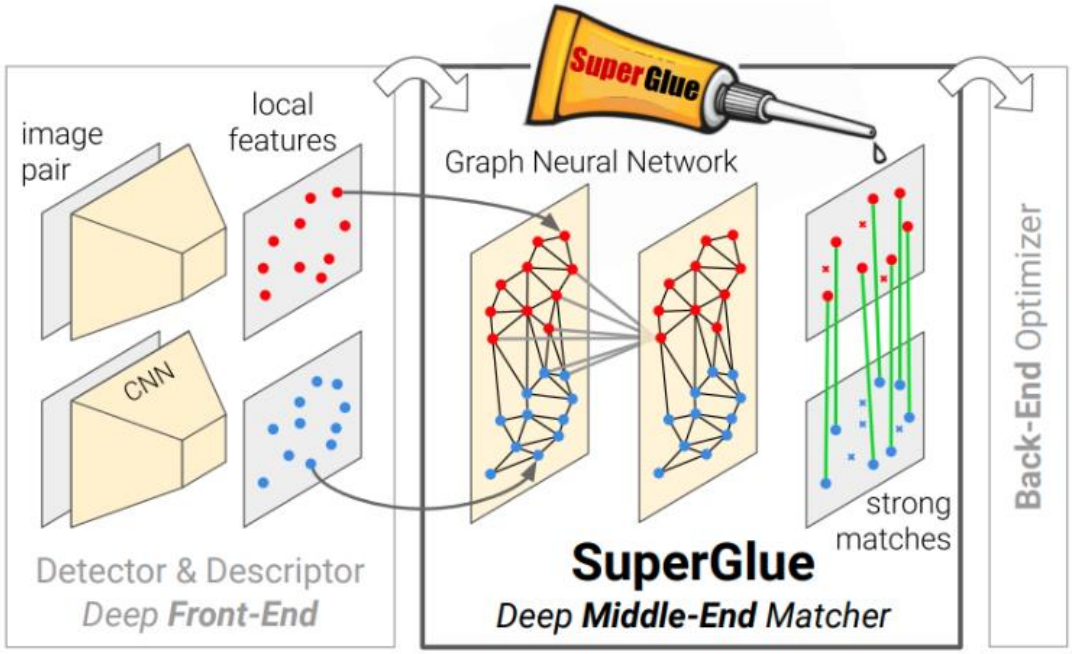
- What is local feature matching
  - Feature extractor (Detector)
    - SIFT (2004), SURF (2006), ORB (2011), SuperPoint (2017), DISK (2020), etc.
  - Feature matcher
    - SIFT (2004), LIFT (2016), SuperGlue (2020), LightGlue (2023), etc.



< Deep learning 을 이용한 local feature matching models 분류 >

Background

- SuperGlue : Learning Feature Matching with Graph Neural Networks
  - Introduction
    - Front-end 와 back-end 사이의 matcher 로써 월등한 성능을 보임
    - Graph Neural network (GNN) 을 사용하여 matching process 를 학습
    - Optimal transport 를 결합하여 matching point 판별



< SuperGlue >

Background

- SuperGlue : Learning Feature Matching with Graph Neural Networks
  - Architecture
    - Attentional graph neural network
      - Keypoint encoder
      - Multiplex graph neural network
        - ✓두 image 의 keypoints 를 complete graph 로 간주
      - Attentional Aggregation
        - ✓Self-edge 일 경우 self-attention 진행
        - ✓Cross-edge 일 경우 cross-attention 진행

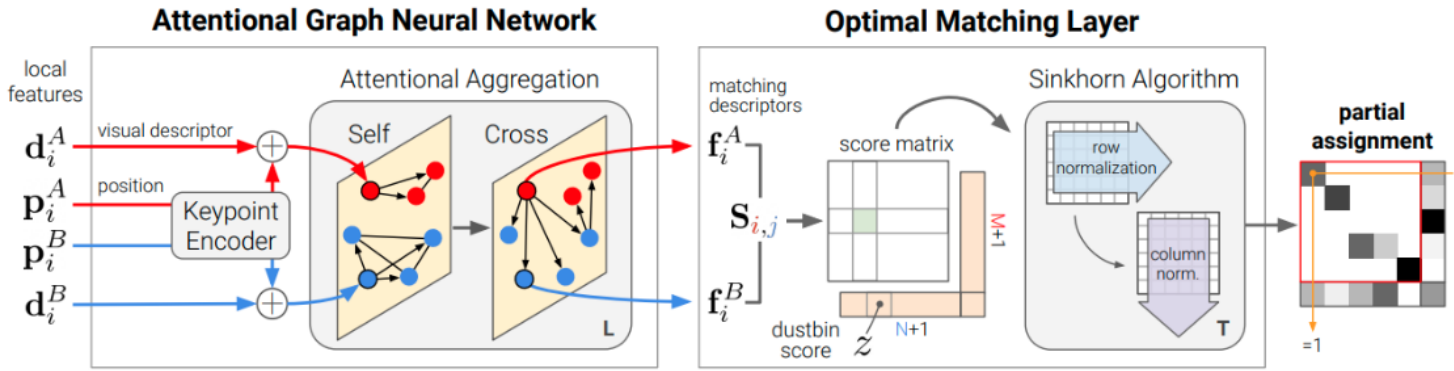
$$ {}^{(0)}\mathbf{x}_i = \mathbf{d}_i + \mathrm{MLP}_{\mathrm{enc}}\left(\mathbf{p}_i\right) $$

$$ {}^{(\ell+1)}\mathbf{x}_i^A = {}^{(\ell)}\mathbf{x}_i^A + \mathrm{MLP}\left(\left[{}^{(\ell)}\mathbf{x}_i^A \,\|\, \mathbf{m}_{\mathcal{E}\to i}\right]\right) $$

$$ \mathbf{m}_{\mathcal{E}\to i} = \sum_{j:(i,j)\in\mathcal{E}} \alpha_{ij}\mathbf{v}_j $$

$\varepsilon \in \{\varepsilon\text{-self}, \varepsilon\text{-cross}\}$

$$ \alpha_{ij} = \mathrm{Softmax}_j\left(\mathbf{q}_i^\top \mathbf{k}_j\right) $$



< SuperGlue 의 architecture >

Background

- SuperGlue : Learning Feature Matching with Graph Neural Networks
  - Architecture
    - Optimal matching layer
      - Score prediction
        - ✓Score 는 match feature 의 similarity

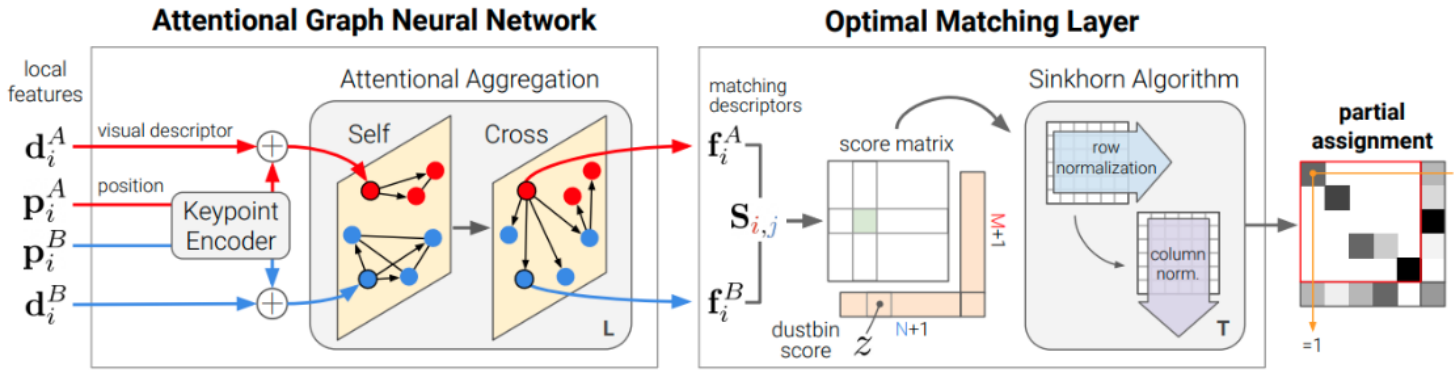
$$\mathbf{S}_{i,j} =< \mathbf{f}_i^A, \mathbf{f}_j^B >, \forall (i,j) \in \mathcal{A} \times \mathcal{B}$$

      - Occlusion and visibility
        - ✓Unmatched keypoints 를 표현할 수 있도록 dustbin 을 추가

$$\bar{\mathbf{S}}_{i,N+1} = \bar{\mathbf{S}}_{M+1,j} = \bar{\mathbf{S}}_{M+1,N+1} = z \in \mathbb{R}$$

      - Sinkhorn algorithm
        - ✓Optimal transport problem
        - ✓Sinkhorn algorithm 의 output 을 통해 partial assignment P 를 반환

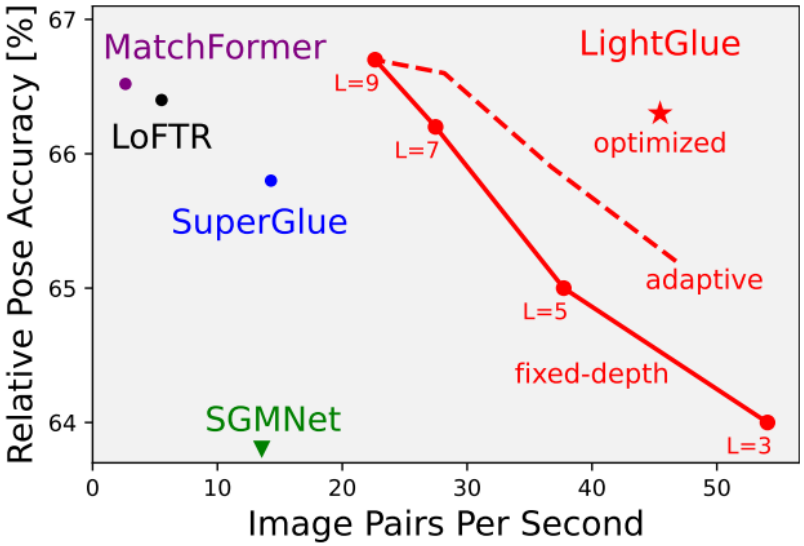
$$\bar{\mathbf{P}}\mathbf{1}_{N+1} = \mathbf{a} \quad \text{and} \quad \bar{\mathbf{P}}^\top \mathbf{1}_{M+1} = \mathbf{b}$$



< SuperGlue 의 architecture >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

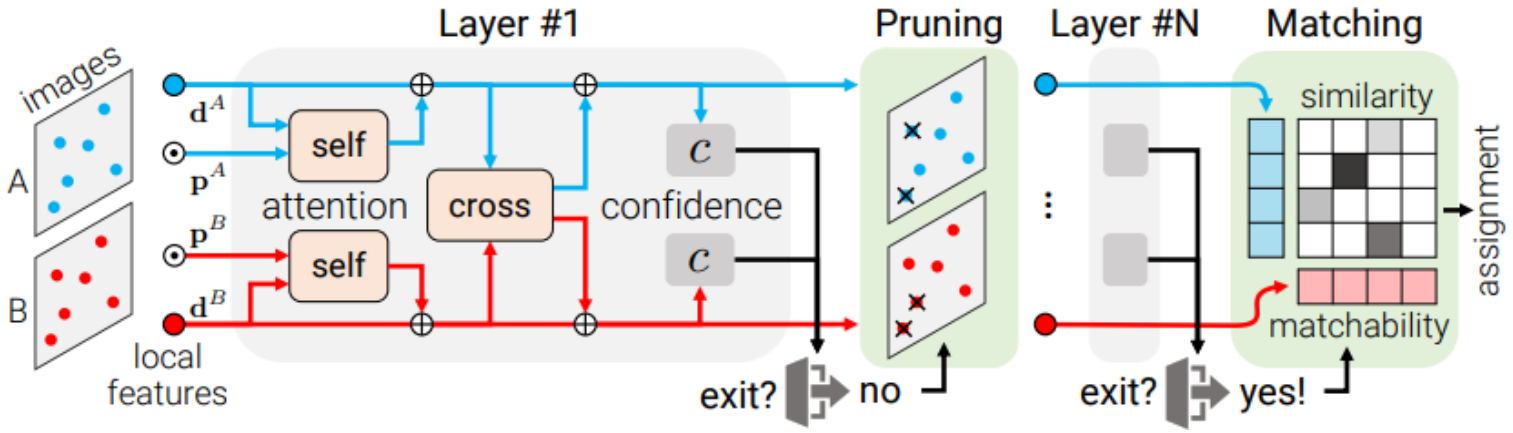
- Introduction
  - 기존의 SuperGlue 방법을 변형
    - Computational block 을 통해 correspondence 집합을 예측
    - Model 이 이를 검사하여 추가계산이 필요한지 여부를 검사
    - 기존의 다른 feature matcher에 비해 높은 성능을 보이며 SOTA 달성



< 여러 feature matcher 들의 속도와 정확도 비교 >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Problem formulation
    - 두 image A, B에 set of local feature 가 존재 $\mathcal{A} := \{1, ..., M\}$ and $\mathcal{B} := \{1, ..., N\}$
    - 각각의 points 는 최대 하나의 point 와 매칭할 수 있음 $\mathcal{M} = \{(i, j)\} \subset \mathcal{A} \times \mathcal{B}$
    - Soft partial assignment 인 $\mathbf{P} \in [0, 1]^{M \times N}$ 를 구하여 correspondences 를 추출



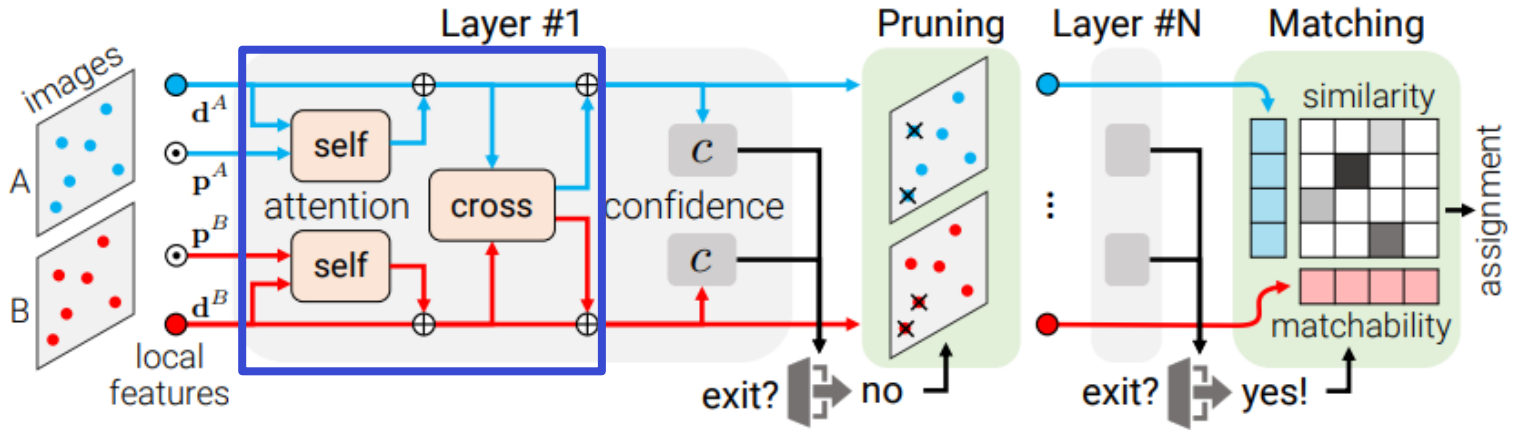
< LightGlue 의 architecture >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Transformer backbone
    - Attention unit
      - Attention unit 은MLP 를 사용하여 상태를 update
      - Self-attention, cross-attention 으로 구성

$$\mathbf{x}_i^I \leftarrow \mathbf{x}_i^I + \mathrm{MLP}\left(\left[\mathbf{x}_i^I \mid \mathbf{m}_i^{I \leftarrow S}\right]\right)$$

$$\mathbf{m}_i^{I \leftarrow S} = \sum_{j \in \mathcal{S}} \underset{k \in \mathcal{S}}{\mathrm{Softmax}}\left(a_{ik}^{IS}\right)_j \mathbf{W}\mathbf{x}_j^S$$



< LightGlue 의 architecture >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Transformer backbone
    - Attention unit

$$\mathbf{x}_i^I \leftarrow \mathbf{x}_i^I + \mathrm{MLP}\left(\left[\mathbf{x}_i^I \mid \mathbf{m}_i^{I \leftarrow S}\right]\right)$$

$$\mathbf{m}_i^{I \leftarrow S} = \sum_{j \in \mathcal{S}} \underset{k \in \mathcal{S}}{\mathrm{Softmax}}\left(a_{ik}^{IS}\right)_j \mathbf{W}\mathbf{x}_j^S$$

$$a_{ij} = \mathbf{q}_i^\top \mathbf{R}(\mathbf{p}_j - \mathbf{p}_i)\, \mathbf{k}_j$$

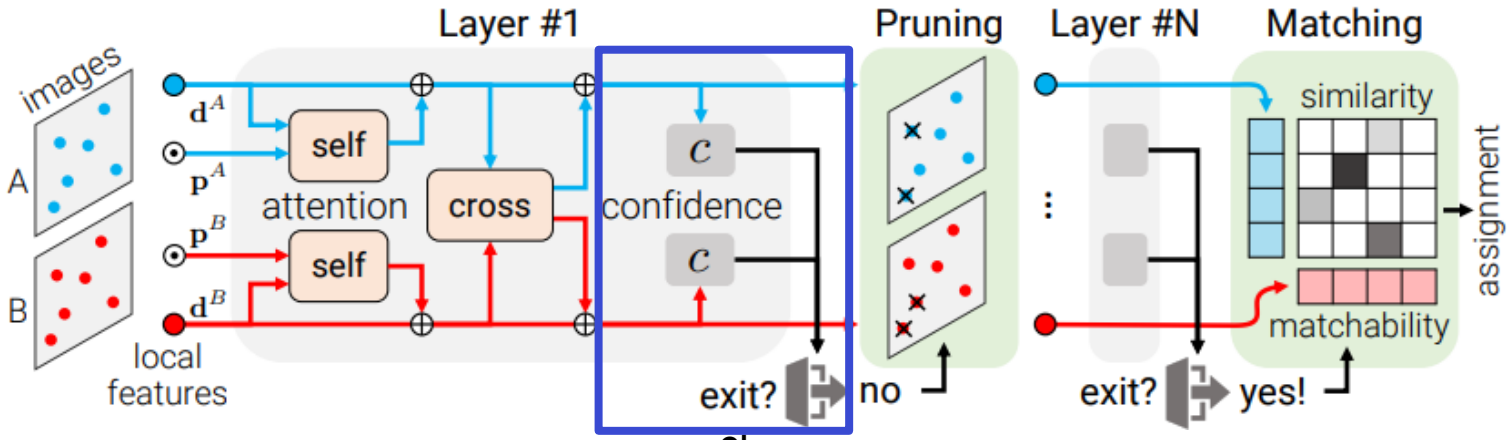
      - Self-attention
        - ✓Rotary position encoding 사용
        - ✓절대적 position 이 아닌 상대적 position 사용
      - Cross-attention
        - ✓query 는 계산하지 않고 key 만 계산
        - ✓Position encoding 도 진행하지 않음
        - ✓ a_{ij}^{IS} 의 값과 a_{ji}^{SI}의 값이 동일하므로 같이 사용
          - Bidirectional attention

$$a_{ij}^{IS} = \mathbf{k}_i^{I\top} \mathbf{k}_j^S \overset{!}{=} a_{ji}^{SI}$$

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Adaptive depth and width
    - 쉬운 image pair 가 input 으로 주어졌을 때, 신뢰할 수 있는 경우가 많음
      - 이 때 초기 layer 의 prediction 과 후기 layer 의 prediction 이 비슷
        - ✓조기 중단 가능
    - Confidence classifier
      - 각 layer 의 끝에서 각 점의 confidence 를 추출

$$c_i = \text{Sigmoid}\left(\text{MLP}(\mathbf{x}_i)\right) \in [0, 1]$$



< LightGlue 의 architecture >

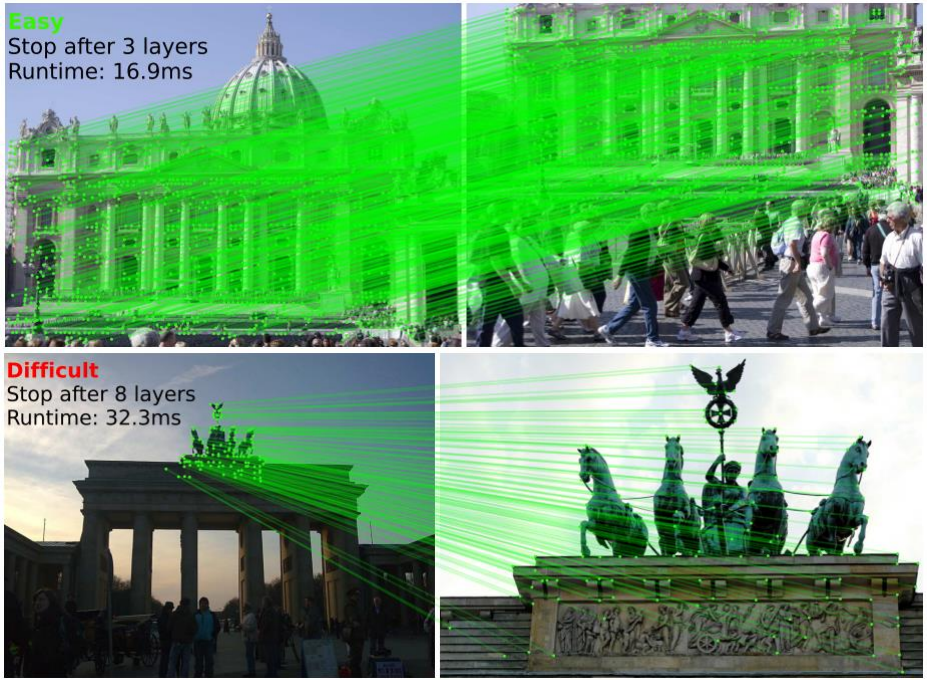
LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Adaptive depth and width
    - Exit criterion
      - 임계값 λ_ℓ, α 에 따라 exit 반환

$$c_i = \text{Sigmoid}\left(\text{MLP}(\mathbf{x}_i)\right) \in [0, 1]$$

$$\text{exit} = \left(\frac{1}{N+M} \sum_{I \in \{A,B\}} \sum_{i \in \mathcal{I}} [\![c_i^I > \lambda_\ell]\!] \right) > \alpha$$

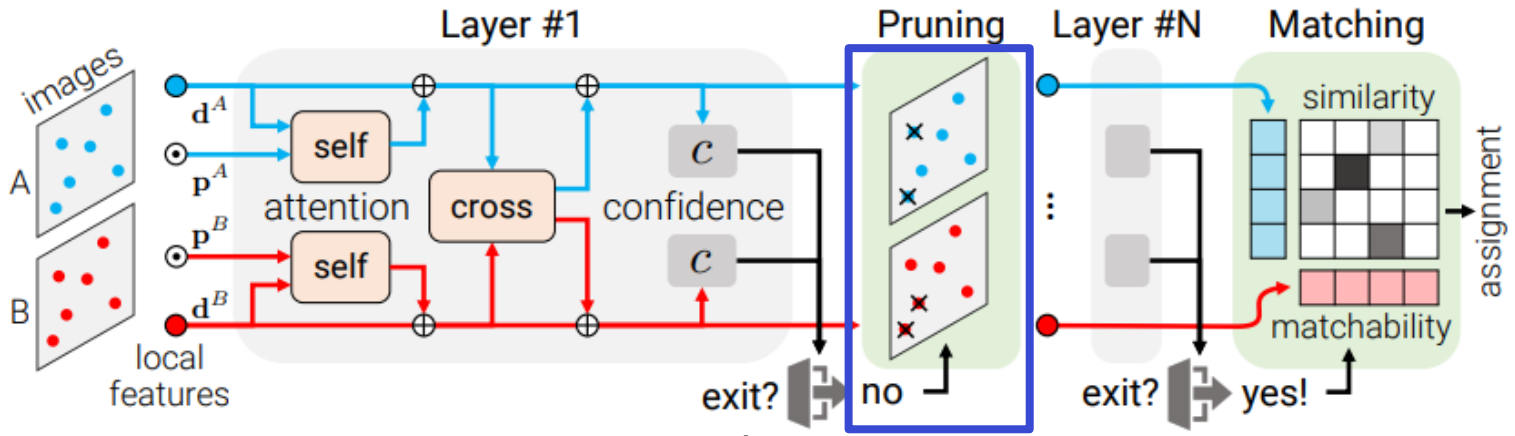
$$\lambda_l = 0.8 + 0.1e^{-4\ell/L}$$



< Depth adaptivity>

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Adaptive depth and width
    - Point pruning
      - Layer 가 지날 수록 일치 하지 않는 keypoints 를 제거



< LightGlue 의 architecture >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Adaptive depth and width
    - Point pruning

Assignment score

✓각 points 의 유사도를 기준으로 score matrix S 를 정의

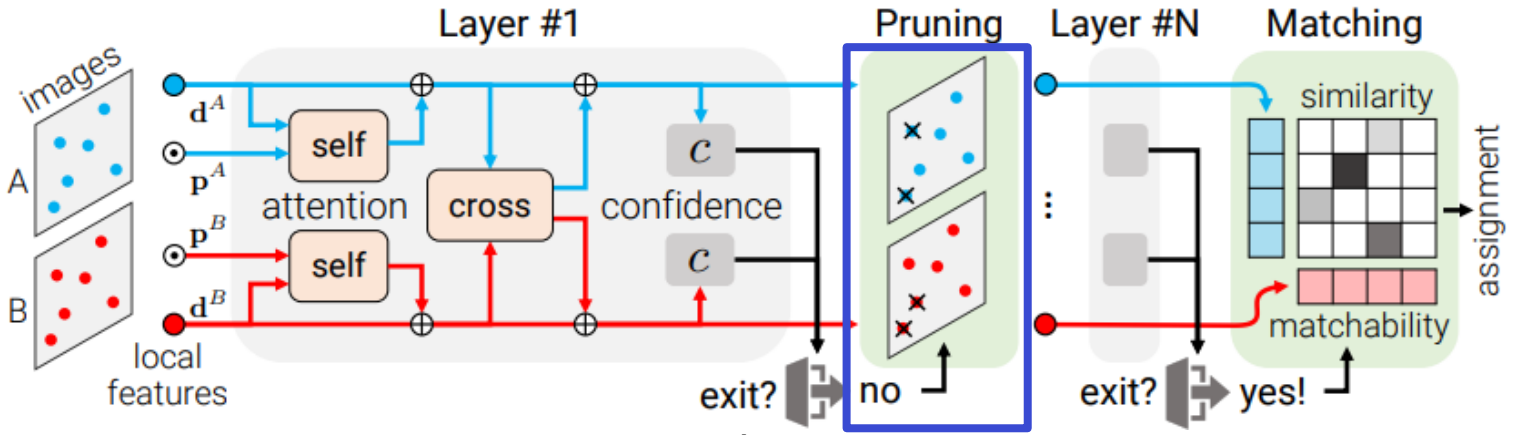
$$\mathbf{S}_{ij} = \text{Linear}\left(\mathbf{x}_i^A\right)^\top \text{Linear}\left(\mathbf{x}_j^B\right) \quad \forall (i,j) \in \mathcal{A} \times \mathcal{B}$$

$$\sigma_i = \text{Sigmoid}\left(\text{Linear}(\mathbf{x}_i)\right) \in [0,1]$$

Correspondences

✓Soft partial assignment matrix P 를 정의

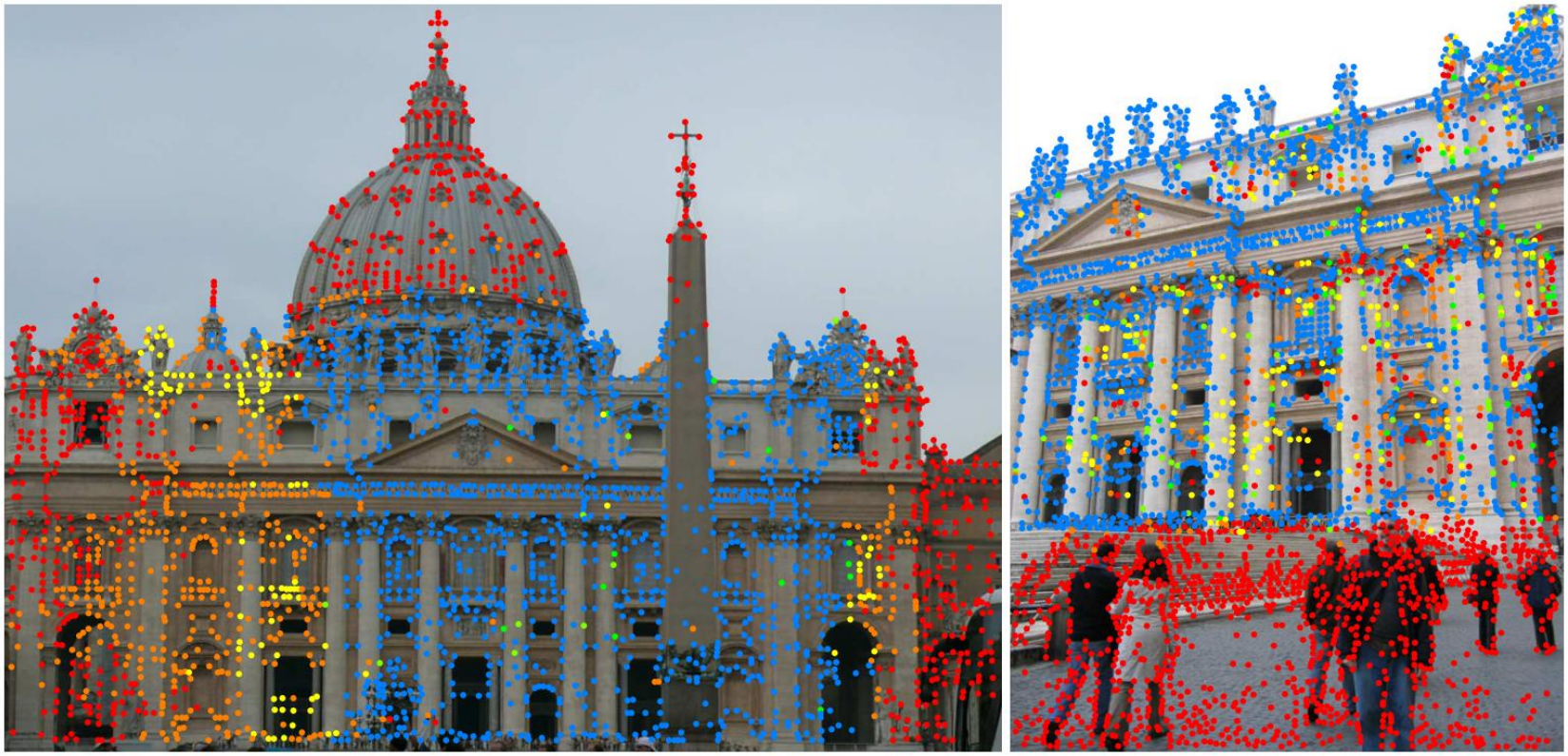
$$\mathbf{P}_{ij} = \sigma_i^A \, \sigma_j^B \underset{k \in \mathcal{A}}{\text{Softmax}}(\mathbf{S}_{kj})_i \underset{k \in \mathcal{B}}{\text{Softmax}}(\mathbf{S}_{ik})_j$$



< LightGlue 의 architecture >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Methodology
  - Adaptive depth and width
    - Point pruning



< Point pruning 의 visualization >

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Supervision
  - Correspondence (similarity)
    - 각 layer ℓ 에서 assignment predicted 의 log-likelihood 를 최소화

$$\text{loss} = -\frac{1}{L}\sum_{\ell}\left(\frac{1}{|\mathcal{M}|}\sum_{(i,j)\in\mathcal{M}}\log {}^{\ell}\mathbf{P}_{ij} + \frac{1}{2|\bar{\mathcal{A}}|}\sum_{i\in\bar{\mathcal{A}}}\log\left(1-{}^{\ell}\sigma_i^A\right) + \frac{1}{2|\bar{\mathcal{B}}|}\sum_{j\in\bar{\mathcal{B}}}\log\left(1-{}^{\ell}\sigma_j^B\right)\right)$$

  - Confidence classifier (matchability)
    - Binary cross-entropy

$$c_i = \text{Sigmoid}\left(\text{MLP}(\mathbf{x}_i)\right) \in [0,1]$$

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Experiment
  - Pose estimation on Megadepth-1800

features + matcher		#matches	P	pose estimation AUC @5°	@10°	@20°	time (ms)
dense	LoFTR	2231	89.8	66.4	79.1	87.6	181
	MatchFormer	2416	91.2	65.2	78.1	87.4	388
	ASPanFormer	4299	94.7	68.0	80.4	88.7	239
SIFT	NN+ratio	160	82.3	48.3	62.2	73.2	5.7
	SGMNet	405	82.5	50.7	66.6	76.5	71.7
	LightGlue	383	84.1	57.0	71.3	81.8	44.3
SuperPoint	NN+mutual	697	49.4	37.7	50.9	62.3	5.6
	SuperGlue	712	93.0	64.8	77.5	86.6	70.0
	SGMNet	**725**	89.8	61.7	74.3	83.4	74.0
	LightGlue	709	**94.5**	**65.5**	**77.8**	**86.9**	44.2

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Experiment
  - Indoor visual localization on InLoc

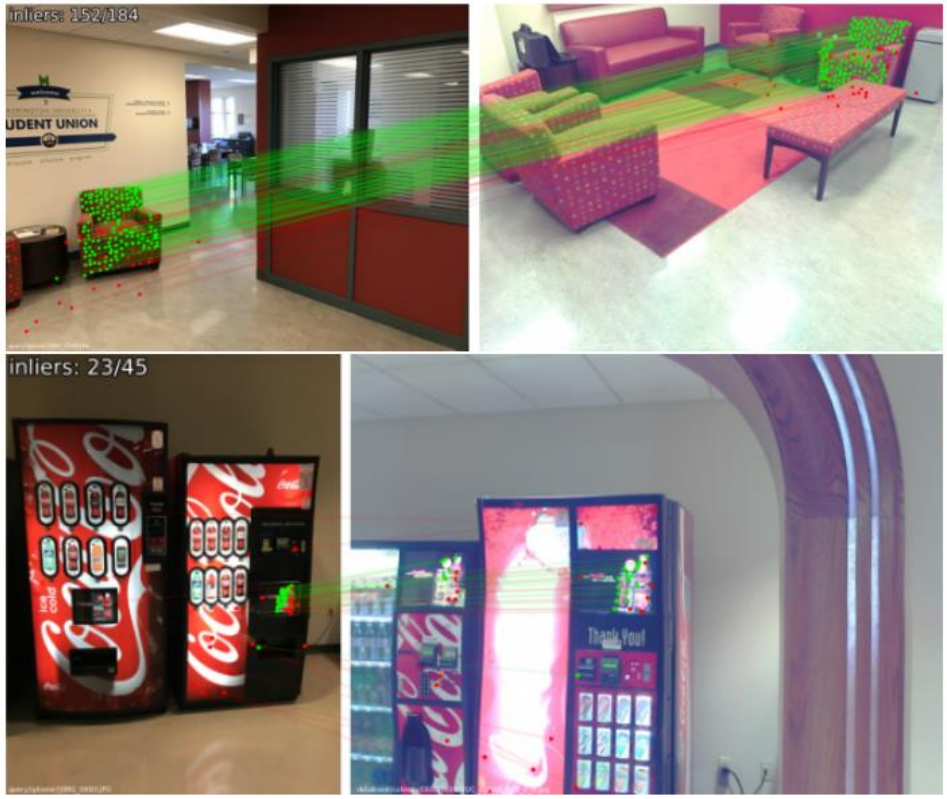
features + matcher	DUC1	DUC2
	(0.25m,10°) / (0.5m,10°) / (1.0m,10°)	
LoFTR	47.5 / 72.2 / 84.8	54.2 / 74.8 / **85.5**
MatchFormer	46.5 / 73.2 / 85.9	**55.7** / 71.8 / 81.7
ASpanFormer	**51.5** / **73.7** / **86.4**	55.0 / 74.0 / 81.7
SP+SuperGlue	47.0 / 69.2 / 79.8	53.4 / **77.1** / 80.9
SP+**LightGlue**	49.0 / 68.2 / 79.3	55.0 / 74.8 / 79.4

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Conclusion
  - Local feature matching 분야에서 graph based-method 를 성공적으로 도입시킨 SuperGlue 를 발전 시킨 network
  - Attention mechanism 을 기반으로 스스로 supervised 가 가능
    - 모든 predict 가 완료됐다고 판단하면 초기 layer 에서 멈춤
    - Matchability 가 낮은 point 는 pruning
  - SuperGlue 보다 빠르고 정확하며 훈련하기 쉬운 model

LightGlue: Local Feature Matching at Light Speed (ICCV 2023)

- Limitation
  - Indoor visual localization on InLoc



감사합니다