



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2023년10월24일

(11) 등록번호 10-2593031

(24) 등록일자 2023년10월19일

(51) 국제특허분류(Int. Cl.)
 G06V 10/77 (2022.01) G06N 3/0464 (2023.01)
 G06N 3/08 (2023.01) G06V 10/48 (2022.01)
 G06V 10/82 (2022.01)

(52) CPC특허분류
 G06V 10/7715 (2023.08)
 G06N 3/0464 (2023.01)

(21) 출원번호 10-2023-0049379

(22) 출원일자 2023년04월14일

심사청구일자 2023년04월14일

(30) 우선권주장
 1020230032445 2023년03월13일 대한민국(KR)

(56) 선행기술조사문헌
 KR1020210057847 A
 CN108920643 A

(73) 특허권자

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

김승룡

서울특별시 성북구 오패산로 46, 103동 1804호(하월곡동, 월곡두산위브아파트)

남지수

서울특별시 성북구 고려대로 29, 402호(안암동2가, SM빌딩)

(뒷면에 계속)

(74) 대리인

송인호, 윤형근, 최관락

전체 청구항 수 : 총 5 항

심사관 : 손윤식

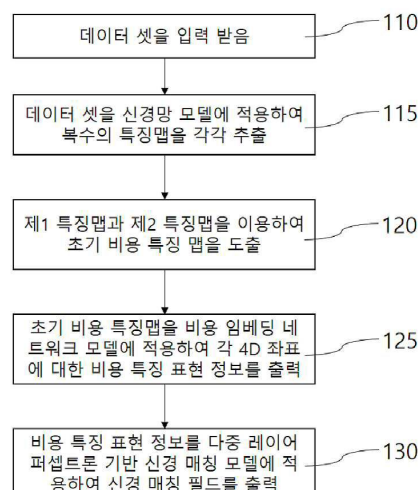
(54) 발명의 명칭 고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치

(57) 요약

고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치가 개시된다. 고정밀 대응을 위한 신경 매칭 표현 방법은, 데이터 셋을 신경망 모델에 각각 적용하여 복수의 특징맵을 추출한 후 상기 복수의 특징맵을 이용하여 원시 비용 특징맵을 도출하는 단계-상기 데이터 셋은 소스 영상과 타겟 영상을 포함함; 상기 원시 비용 특징맵을 비용 임베

딩 네트워크 모델에 적용하여 비용 특징 표현 정보를 출력하는 단계-상기 비용 특징 표현 정보 $\phi(C, p)$ 는 원시 비용 특징맵에서 추출된 4D 좌표 p 의 비용 특징 벡터를 포함함; 및 상기 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06N 3/08 (2023.01)
G06V 10/48 (2023.08)
G06V 10/82 (2022.01)

(72) 발명자

홍성환

서울특별시 강북구 한천로109길 57-56, B01호 (변동)

조석주

서울특별시 성북구 성북로16가길 7-6(성북동)

홍수성

서울특별시 영등포구 의사당대로 127, 101-1502호
(여의도동, 롯데캐슬엠펙라이어)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711159611
과제번호	2020-0-01819-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	ICT명품인재양성(고려대학교)
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2022.01.01 ~ 2022.12.31

공지예외적용 : 있음

명세서

청구범위

청구항 1

데이터 셋을 신경망 모델에 각각 적용하여 복수의 특징맵을 추출한 후 상기 복수의 특징맵을 이용하여 원시 비용 특징맵을 도출하는 단계-상기 데이터 셋은 소스 영상과 타겟 영상을 포함함;

상기 원시 비용 특징맵을 비용 임베딩 네트워크 모델에 적용하여 비용 특징 표현 정보를 출력하는 단계-상기 비

용 특징 표현 정보 $\phi(C, \mathbf{p})$ 는 원시 비용 특징맵에서 추출된 4D 좌표 $\mathbf{p}$ 의 비용 특징 벡터를 포함함;

상기 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력하는 단계를 포함하되,

상기 4D 좌표는 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 이며, $\mathbf{x} \in [0, H_s) \times [0, W_s)$ 이고, $\mathbf{y} \in [0, H_t) \times [0, W_t)$ 이되, 상기 H_s 와 W_s 는 상기 소스 영상의 해상도를 나타내고, H_t 와 W_t 는 상기 타겟 영상의 해상도를 나타내는 것을 특징으로 하는 고정밀 대응을 위한 신경 매칭 표현 방법.

청구항 2

제1 항에 있어서,

상기 신경 매칭 필드는 하기 수학식을 이용하여 도출되는 것을 특징으로 하는 고정밀 대응을 위한 신경 매칭 표현 방법.

$$M(\mathbf{p}) = f_{\theta}(\gamma(\mathbf{p}), \phi(C, \mathbf{p})) \in [0, 1],$$

여기서, $\gamma(\mathbf{p})$ 는 $\mathbf{p}$ 의 부호화된 위치를 나타내고, 상기 f_{θ} 는 매개변수 θ 를 가지는 상기 다중 레이어 퍼셉트론 모델의 매칭 함수를 나타냄.

청구항 3

제1 항에 있어서,

상기 다중 레이어 퍼셉트론 기반 신경 매칭 모델의 손실 함수는 하기 수학식으로 계산되는 것을 특징으로 하는 고정밀 대응을 위한 신경 매칭 표현 방법.

$$\mathcal{L}_{\text{total}} = \lambda_f \mathcal{L}_f + \lambda_c \mathcal{L}_c$$

$$\mathcal{L}_f = - \sum_{k=1}^K M_k^* \log(M(\mathbf{p}_k)).$$

여기서, λ_f 와 λ_c 는 각각 가중치를 나타내고,

이되,

$\mathbf{p}_k$ 는 쿼리 샘플을 나타내고, M_k^* 는 실제 신경 매칭 필드를 나타내며, $\mathcal{L}_c = \|F_{gt} - F_{pred}\|_2$ 이며, F_{pred} 는 채널별 평균 합산 비용 특징 블록 $V = \text{avgpool}(\phi(C, \mathbf{p})) \in \mathbb{R}^{H_s \times W_s \times H_t \times W_t}$ 에 소프트맥스 함수를 적용하여 계산된 결과이며, F_{gt} 는 실측값인 것이고, $\|\cdot\|$ 는 l_2 norms 을 나타냄.

청구항 4

제1 항에 있어서,

상기 신경 매칭 필드를 이용하여 상기 소스 영상과 상기 타겟 영상의 대응 필드가 결정되는 것을 특징으로 하는 고정밀 대응을 위한 신경 매칭 표현 방법.

청구항 5

적어도 하나의 명령어를 저장하는 메모리; 및

상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되,

상기 프로세서에 의해 실행된 명령어는 각각,

데이터 셋을 신경망 모델에 각각 적용하여 복수의 특징맵을 추출한 후 상기 복수의 특징맵을 이용하여 원시 비용 특징맵을 도출하는 단계-상기 데이터 셋은 소스 영상과 타겟 영상을 포함함;

상기 원시 비용 특징맵을 비용 임베딩 네트워크 모델에 적용하여 비용 특징 표현 정보를 출력하는 단계-상기 비

용 특징 표현 정보 $\phi(C, \mathbf{p})$ 는 원시 비용 특징맵에서 추출된 4D 좌표 $\mathbf{p}$ 의 비용 특징 벡터를 포함함; 및

상기 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력하는 단계를 수행하되,

상기 4D 좌표는 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 이며, $\mathbf{x} \in [0, H_s) \times [0, W_s)$ 이고,

$\mathbf{y} \in [0, H_t) \times [0, W_t)$ 이되, 상기 H_s 와 W_s 는 상기 소스 영상의 해상도를 나타내고, H_t

와 W_t 는 상기 타겟 영상의 해상도를 나타내는 것을 특징으로 하는 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명은 고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치에 관한 것이다.

배경 기술

[0003] 초기 연구들은 수작업으로 제작된 특징 설명자를 제안하여 특징 추출 단계에 초점이 맞춰져 있었다. 먼저 관심 영역을 감지하고, 해당 특징 설명을 추출한 후 대응의 초기 집합을 설정하였다. 이러한 작업은 전통적인 수작업 특징 설명자를 기반으로 하지만, 높은 수준의 의미를 포착하는데 제한적인 능력을 보인다.

[0004] 이러한 문제를 해결하기 위해 CNN은 처음에는 수작업으로 만든 특징을 심층 특징으로 대체하여 종단간 학습으로

빠르게 수렴하는 강력한 표현 덕분에 패러다임 전환을 이루었다.

- [0005] CNN을 사용하여 추출된 심층 특징맵과 추출된 특징 맵 간에 계산된 상관 관계 맵을 기반을 중단간 기하학적 매칭 네트워크가 제안되었다. 그 이후로 설명자간의 유사성 쌍을 모두 포함하는 상관 관계 맵을 사용하는 방법이 많은 매칭 네트워크에서 주로 사용되었다.
- [0006] 그러나, 모든 쌍별 유사성을 모두 계산하고 저장하는 것은 입력 공간의 크기에 대한 2차 메모리와 계산 복잡성을 필요로 할 뿐만 아니라, 해상도 증가로 인해 이들을 모두 계산하는 것은 불가능하며, 필연적으로 낮은 분해능에서 정의된 상관 관계 맵을 활용하게 되었다.
- [0007] 다른 방법으로 NC-Net이 있는데, 이웃 컨센서스를 탐색하여 공간적으로 일관된 일치를 식별하기 위해 4D 컨볼루션 사용을 제안하였으며, DHPF은 확률론적 Hough 매칭(PHM)을 적용하여 일치점을 찾는 방식이다.
- [0008] 또한, CHM은 6D 상관 관계 맵을 집계하기 위해 고차원 컨볼루션 커널을 사용하여 PHM을 확장하나, 이들은 필연적으로 예측된 대응 필드를 원래 영상 레벨 해상도로 업샘플링해야 하므로 정밀도가 떨어지는 단점이 있다.
- [0009] 최근 암시적 신경 표현(INR: implicit neural representation)이 큰 관심을 받고 있다. 이는 좌표 기반 표현으로, 신경망에 의해 매개변수화된 연속적이고 미분 가능한 신호 표현이다. 암시적 신경 표현(INR)은 공간 분해능에 결합되지 않기 때문에 입력 신호를 공간 분해능에 직교하는 매개 변수화하기 위한 메모리 사양이 필요하다.

발명의 내용

해결하려는 과제

- [0011] 본 발명은 고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치가 제공된다.
- [0012] 또한, 본 발명은 원래 영상의 해상도에서 정의된 심층 완전 연결 네트워크를 통해 고차원 4D 매칭 필드를 암시적으로 나타낼 수 있는 고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치를 제공하기 위한 것이다.

과제의 해결 수단

- [0014] 본 발명의 일 측면에 따르면, 고정밀 대응을 위한 신경 매칭 표현 방법이 제공된다.
- [0015] 본 발명의 일 실시예에 따르면, 데이터 셋을 신경망 모델에 각각 적용하여 복수의 특징맵을 추출한 후 상기 복수의 특징맵을 이용하여 원시 비용 특징맵을 도출하는 단계-상기 데이터 셋은 소스 영상과 타겟 영상을 포함함; 상기 원시 비용 특징맵을 비용 임베딩 네트워크 모델에 적용하여 비용 특징 표현 정보를 출력하는 단계-상기 비

용 특징 표현 정보 $\phi(C, \mathbf{p})$ 는 원시 비용 특징맵에서 추출된 4D 좌표 $\mathbf{p}$ 의 비용 특징 벡터를 포함함; 및 상기 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력하는 단계를 포함하되, 상기 4D 좌표는 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 이며, $\mathbf{x} \in [0, H_s) \times [0, W_s)$ 이고, $\mathbf{y} \in [0, H_t) \times [0, W_t)$ 이되, 상기 H_s 와 W_s 는 상기 소스 영상의 해상도를 나타내고, H_t 와 W_t 는 상기 타겟 영상의 해상도를 나타내는 것을 특징으로 하는 고정밀 대응을 위한 신경 매칭 표현 방법이 제공될 수 있다.

- [0016] 상기 신경 매칭 필드는 하기 수학식을 이용하여 도출되되,

$$M(\mathbf{p}) = f_{\theta}(\gamma(\mathbf{p}), \phi(C, \mathbf{p})) \in [0, 1],$$

- [0017] 여기서, $\gamma(\mathbf{p})$ 는 $\mathbf{p}$ 의 부호화된 위치를 나

타내고, 상기 f_{θ} 는 매개변수 θ 를 가지는 상기 다중 레이어 퍼셉트론 모델의 매칭 함수를 나타낸다.

[0018] 상기 다중 레이어 퍼셉트론 기반 신경 매칭 모델의 손실 함수는 하기 수학식으로 계산되되,

[0019]
$$\mathcal{L}_{\text{total}} = \lambda_f \mathcal{L}_f + \lambda_c \mathcal{L}_c$$

[0020] 여기서, λ_f 와 λ_c 는 각각 가중치를 나타내고,
$$\mathcal{L}_f = - \sum_{k=1}^K M_k^* \log(M(\mathbf{p}_k)),$$
 이되,

$\mathbf{p}_k$ 는 쿼리 샘플을 나타내고, M_k^* 는 실제 신경 매칭 필드를 나타내며,
 $\mathcal{L}_c = \|F_{\text{gt}} - F_{\text{pred}}\|_2$ 이며, F_{pred} 는 채널별 평균 합산 비용 특징 블록
 $V = \text{avgpool}(\phi(C, \mathbf{p})) \in \mathbb{R}^{H_s \times W_s \times H_t \times W_t}$ 에 소프트맥스 함수를 적용하여
 계산된 결과이며, F_{gt} 는 실측값인 것이고, $\|\cdot\|$ 는 l_2 norms 을 나타낸다.

[0021] 상기 신경 매칭 필드를 이용하여 상기 소스 영상과 상기 타겟 영상의 대응 필드가 결정될 수 있다.

[0023] 본 발명의 다른 측면에 따르면, 고정밀 대응을 위한 신경 매칭 표현 방법을 수행하는 장치가 제공된다.

[0024] 본 발명의 일 실시예에 따르면, 적어도 하나의 명령어를 저장하는 메모리; 및 상기 메모리에 저장된 명령어를 실행하는 프로세서를 포함하되, 상기 프로세서에 의해 실행된 명령어는 각각, 데이터 셋을 신경망 모델에 각각 적용하여 복수의 특징맵을 추출한 후 상기 복수의 특징맵을 이용하여 원시 비용 특징맵을 도출하는 단계-상기 데이터 셋은 소스 영상과 타겟 영상을 포함함; 상기 원시 비용 특징맵을 비용 임베딩 네트워크 모델에 적용하

여 비용 특징 표현 정보를 출력하는 단계-상기 비용 특징 표현 정보 $\phi(C, \mathbf{p})$ 는 원시 비용 특징맵에서 추출된 4D 좌표 $\mathbf{p}$ 의 비용 특징 벡터를 포함함; 및 상기 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력하는 단계를 수행하되, 상기 4D 좌표는 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 이며, $\mathbf{x} \in [0, H_s) \times [0, W_s)$ 이고, $\mathbf{y} \in [0, H_t) \times [0, W_t)$ 이되, 상기 H_s 와 W_s 는 상기 소스 영상의 해상도를 나타내고, H_t 와 W_t 는 상기 타겟 영상의 해상도를 나타내는 것을 특징으로 하는 컴퓨팅 장치가 제공될 수 있다.

발명의 효과

[0026] 본 발명의 일 실시예에 따른 고정밀 대응을 위한 신경 매칭 표현 방법 및 그 장치를 제공함으로써, 원래 영상의 해상도에서 정의된 심층 완전 연결 네트워크를 통해 고차원 4D 매칭 필드를 암시적으로 나타낼 수 있다.

[0027] 이를 통해, 본 발명은 종래에 비해 높은 정밀도 대응을 효과적으로 찾을 수 있다.

도면의 간단한 설명

[0029] 도 1은 본 발명의 일 실시예에 따른 고정밀 대응을 위한 신경 매칭 표현 방법을 나타낸 순서도.

도 2는 본 발명의 일 실시예에 따른 모델 구조를 도시한 도면.

도 3은 본 발명의 일 실시예에 따른 훈련 과정을 설명하기 위해 도면.

도 4는 본 발명의 일 실시예에 따른 추론 과정을 설명하기 위해 도시한 도면.

도 5는 본 발명의 일 실시예에 따른 서로 다른 반복 횟수에 따른 플로우 맵을 예시한 도면.

도 6은 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 도면.

발명을 실시하기 위한 구체적인 내용

[0030] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다. 또한, 명세서에 기재된 "...부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

[0031] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세히 설명한다.

[0033] 도 1은 본 발명의 일 실시예에 따른 고정밀 대응을 위한 신경 매칭 표현 방법을 나타낸 순서도이다.

[0034] 단계 110에서 컴퓨팅 장치(100)는 데이터 셋을 입력받는다.

[0035] 여기서, 데이터 셋은 소스 영상과 타겟 영상을 포함할 수 있다. 소스 영상을 $I_s \in \mathbb{R}^{H_s \times W_s}$ 와 같이

병행 표기하며, 타겟 영상은 $I_t \in \mathbb{R}^{H_t \times W_t}$ 로 표기하기로 한다.

[0036] 단계 115에서 컴퓨팅 장치(100)는 데이터 셋을 신경망 모델에 적용하여 복수의 특징맵을 각각 추출한다. 즉, 컴퓨팅 장치(100)는 소스 영상과 타겟 영상을 신경망 모델에 적용하여 제1 특징맵과 제2 특징맵을 각각 추출한다. 이하에서는 이해와 설명의 편의를 도모하기 위해, 소스 영상의 특징맵을 제1 특징맵이라 칭하기로 하며, 타겟 영상의 특징맵을 제2 특징맵이라 칭하여 설명하기로 한다.

[0037] 또한, 신경망 모델을 통해 특징맵을 추출하는 방법 자체는 공지된 사항이므로 이에 대한 별도의 설명은 생략하기로 한다.

[0038] 단계 120에서 컴퓨팅 장치(100)는 제1 특징맵과 제2 특징맵을 이용하여 초기 비용 특징 맵을 도출한다.

[0039] 단계 125에서 컴퓨팅 장치(100)는 초기 비용 특징맵을 비용 임베딩 네트워크 모델에 적용하여 각 4D 좌표에 대한 비용 특징 표현 정보를 출력한다.

[0040] 여기서, 비용 임베딩 네트워크 모델은 컨볼루션과 셀프 어텐션 레이어로 구성될 수 있다. 해당 비용 임베딩 네트워크 모델은 도 2의 220에 도시된 바와 같이, 초기 비용 특징맵에서 추출된 4D 좌표 p의 특징 벡터를 포함할 수 있다.

[0041] 이에 대해서는 하기의 설명에 의해 보다 명확하게 이해될 것이다.

[0042] 단계 130에서 컴퓨팅 장치(100)는 비용 특징 표현 정보를 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 신경 매칭 필드를 출력한다. 신경 매칭 필드를 이용하여 소스 영상과 타겟 영상의 대응 필드가 결정될 수 있다.

[0043] 이에 대해 보다 상세히 설명하기로 한다.

[0044] 본 발명의 일 실시예에 따른 컴퓨팅 장치(100)는 소스 영상과 타겟 영상 쌍이 $I_s \in \mathbb{R}^{H_s \times W_s}$ 와

$I_t \in \mathbb{R}^{H_t \times W_t}$ 로 주어지는 경우, 소스 영상 해상도에서 각 픽셀 x에 대해 정의되는 조밀한 대응 필드

$I_t(x) \approx I_s(x + F(x))$ 를 만족시키기 위해 I_s 를 $F(x)$ 를 찾는 것을 목표로 한다. 따라서,

I_t 로 향하도록 이동시킨다.

[0045] 이를 위해, 소스 영상과 타겟 영상을 각각 학습된 신경망 모델에 적용하여 특징맵을 추출한 다음 코사인 거리를 사용하여 전체 쌍별 유사성 점수(이하, 초기 비용 특징 맵이라 칭하기로 함)를 각각 도출한다. 이를 수학식으로 나타내면 수학식 1과 같다.

수학식 1

$$C(\mathbf{x}, \mathbf{y}) = \frac{D_s(\mathbf{x}) \cdot D_t(\mathbf{y})}{\|D_s(\mathbf{x})\| \|D_t(\mathbf{y})\|}$$

여기서, 여기서, $\mathbf{x} \in [0, h_s) \times [0, w_s)$ 이고, $\mathbf{y} \in [0, h_t) \times [0, w_t)$ 이며,

$\|\cdot\|$ 는 l_2 norms 을 나타낸다. 또한, $D_s(\mathbf{x})$ 는 제1 특징맵을 나타내고, $D_t(\mathbf{y})$ 는 제1 특징맵을 각각 나타낸다.

컨볼루션 신경망 모델의 심층 계층에서 소스 영상과 타겟 영상 각각의 특징맵을 추출하여 높은 수준의 의미론적 불변성을 갖도록 하여 제1 특징맵과 제2 특징맵의 공간 해상도가 감소된다(즉, $h < \bar{H}$ 및 $w < W$).

이로 인해, $C(\mathbf{x}, \mathbf{y})$ 를 사용하여 대응 관계를 추론하면, 불가피하게 저해상도 대응 맵이 생성되며, 이로 인해 추가적으로 고해상도 맵으로 보간을 위한 후처리를 필요로 한다.

이에, 본 발명의 일 실시예에서는 별도의 사후 처리 없이 임의의 척도에서 고정밀 대응을 추론하기 위해 고차원 4D 매칭 필드를 암시적으로 나타낸 고정밀 대응을 위한 신경 매칭 필드를 이용하고자 한다.

소스 영상 해상도에서 정의된 4D 좌표를 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 로 표기하기로 한다. 여기서, $\mathbf{x} \in [0, H_s) \times [0, W_s)$ 이고, $\mathbf{y} \in [0, H_t) \times [0, W_t)$ 이다.

신경 매칭 필드는 수학식 2와 같이 계산될 수 있다.

수학식 2

$$M(\mathbf{p}) = f_\theta(\gamma(\mathbf{p}), \phi(C, \mathbf{p})) \in [0, 1],$$

여기서, $\gamma(\mathbf{p})$ 는 $\mathbf{p}$ 의 부호화된 위치를 나타내고, $\phi(C, \mathbf{p})$ 는 초기 비용 특징맵 C 에서 추출된 $\mathbf{p}$ 의 비용 특징 벡터를 나타낸다.

도 2의 230에서 보여지는 바와 같이, 본 발명의 일 실시예에 따르면, 다중 레이어 퍼셉트론 기반 신경 매칭 모델은 파라미터 θ 를 가지는 연속 함수(매칭 함수) f_θ 로 공식화될 수 있다.

매칭 함수 f_θ 의 설계 가능한 방법은 두 입력 $\gamma(\mathbf{p})$ 와 $\phi(C, \mathbf{p})$ 를 연결한 후 완전 연결된 네트워크(즉, 다중 레이어 퍼셉트론 기반 신경 매칭 모델)을 통해 전달하는 것이다. 그러나, 이 방법은 메모리 집약적인 배치 정규화 작업을 강요할 수 있다.

[0057] 따라서, 본 발명의 일 실시예에서는 $\phi(C, \mathbf{p})$ 을 완전히 연결된 각 블록의 입력 특징에 추가하는 것이다. 또한, 4D 신경 매칭 필드의 값이 0에서 1 사이의 범위에 있도록 보정하기 위해 비유입 임베딩 모델의 끝에 시그모이드 함수를 사용할 수도 있다.

[0058] 이해와 설명의 편의를 도모하기 위해, 신경 매칭 필드(NeMF)에 대해 간략하게 설명하기로 한다.

[0059] 신경 매칭 필드(NeRF: Neural Radiance Field)는 $f_{\omega} : (\mathbf{o}, \mathbf{d}) \rightarrow (\sigma, \mathbf{c})$ 와 같이 3D 위치 $\mathbf{o} = (\bar{x}, y, z)$ 및 2D 보기 방향 $\mathbf{d} = (\theta, \phi)$ 를 입력으로 취하는 볼륨 밀도 σ 및 RGB 색상 값 $\mathbf{c} = (r, g, b)$ 를 계산하는 매개변수 ω 가 있는 연속 함수 f_{ω} 로 정의될 수 있다.

[0060] 낮은 차원의 입력 $\mathbf{x}$ 및 $\mathbf{d}$ 를 신경망을 통과하기 전에 더 높은 차원의 기능에 매핑하면 고주파 변동을 더 잘 표현할 수 있다. $\gamma(\cdot)$ 를 인코딩 함수로, L 을 주파수 옥타브의 수로 나타내면, 수학식 3과 같다.

수학식 3

$$\gamma(t) = [\sin(2^0 t \pi), \cos(2^0 t \pi), \dots, \sin(2^L t \pi), \cos(2^L t \pi)]$$

[0061]

[0062] 여기서, 연속 함수 $f_{\omega} : \mathbb{R}^{L_{\mathbf{o}}} \times \mathbb{R}^{L_{\mathbf{d}}} \rightarrow \mathbb{R}^+ \times \mathbb{R}^3$ 는 수학식 4와 같이 나타낼 수 있다.

수학식 4

$$\{\sigma, \mathbf{c}\} = f_{\omega}(\gamma(\mathbf{o}), \gamma(\mathbf{d}))$$

[0063]

[0064] 여기서, $L_{\mathbf{o}}$ 와 $L_{\mathbf{d}}$ 는 각각 인코딩된 좌표 $\mathbf{o}$ 와 보는 방향 $\mathbf{d}$ 의 출력 차원을 나타낸다. 연속 함수 f_{ω} 는 완전 연결된 심층 네트워크로 공식화될 수 있다. 이러한 암시적 신경 표현은 연속 함수를 사용하기 위한 공간 분해능과 결합되지 않으므로, 공간 분해능과 무관하게 신호를 매개변수화하는 데 필요한 메모리 소비를 만든다.

[0065] 본 발명은 이러한 암시적 신경 표현에 착안하여, 4D 좌표 및 초기 원시 비유입 특징맵을 비유입 임베딩 네트워크 모델에 적용하여 비유입 특징 표현 정보를 생성하고, 이를 기반으로 완전 연결된 다중 레이어 퍼셉트론 기반 신경 매칭 모델에 적용하여 비유입 필드를 신경 매칭 필드로 도출하고, 이를 기반으로 대응점을 찾을 수 있다.

[0066] 다시 정리하면, 컴퓨팅 장치(100)는 한쌍의 입력 영상이 주어지면, CNN을 통해 특징맵을 추출한 후 저해상도 초기 원시 비유입 특징맵을 계산할 수 있다. 이어, 컨볼루션과 트랜스포머로 구성된 비유입 임베딩 네트워크에 초기 원시 비유입 특징맵을 공급하여 인코딩된 비유입 특징 표현 정보를 생성한 후 해당 비유입 특징 표현 정보와 좌표를 입력으로 완전 연결된 네트워크인 다중 레이어 퍼셉트론 기반 신경 매칭 모델을 통해 디코딩하여 신경 매칭 필드를 도출할 수 있다.

[0067] 본 발명의 일 실시예에서는 비유입 임베딩 네트워크 모델의 셀프 어텐션 계층의 전역 수용 필드를 통해 초기 원시 비유입 특징맵을 추가로 포함할 수 있다. 이와 같은 경우, 비유입 특징맵의 모든 픽셀에서 명시적으로 인코딩되지만

귀납적 편향, 즉 컨볼루션에 의한 동등성 또는 상대적 위치 편향을 부여하는 작업이 없는 경우 오류가 있는 표현이 생성될 수 있다. 따라서, 비용 임베딩 네트워크 모델은 트랜스포머를 컨볼루션 연산자와 결합하여 유도 바이어스의 부족을 보상하고 로컬 컨텍스트를 캡슐화하고 셀프-어텐션을 통해 모든 픽셀에 전달함으로써 일치하는 신호의 로컬 및 글로벌 통합을 허용할 수 있다.

[0068] 즉, 매칭 함수 f_{θ} 에 일치하는 비용을 제공하기 전에 4D 원시 비용 볼륨 C는 여전히 저해상도에서 정의되는 K 채널과 함께 5D 비용 특징 볼륨 $C' \in \mathbb{R}^{H_s \times W_s \times H_t \times W_t \times K}$ 에 포함될 수 있다.

[0069] 그런 다음 비용 특징 벡터 $\phi(C, p) \in \mathbb{R}^{K \times 1}$ 에 대한 쿼리 포인트 p에 대해 4중 선형 보간법 $\phi(C)$ 을 사용할 수 있다.

[0070] 도 3을 참조하여 훈련 방법에 대해 설명하기로 한다.

[0071] 입력 영상 쌍 사이의 실측 키-포인트 쌍 $\{x, x'\}$ 은 다음과 같은 방법으로 사용한다. 즉, 쿼리 포인트 $p = [x, y]$ 가 실측 대응으로 분류되면, 네트워크 출력은 1이 되고, 그렇지 않으면 0이 되도록 권장된다.

[0072] 본 발명의 일 실시예에서는 이를 분류 문제로 공식화하고 교차 엔트로피 손실을 적용하여 타겟 영상에서 샘플된 네거티브 키포인트 y(여기서, $y \neq x'$) 중에서 소스 영상의 쿼리 x에 대한 대응(즉, 실측 x')의 정확성을 예측하는 방법을 학습할 수 있다.

[0073] 본 발명의 일 실시예에서는 균일한 분포에서 무작위 샘플링을 네거티브 샘플로 채택할 수 있다. 공식적으로 손실 함수는 수학적 5와 같이 정의될 수 있다.

수학적 5

$$\mathcal{L}_f = - \sum_{k=1}^K M_k^* \log(M(p_k)).$$

[0074]

[0075] 여기서, p_k 는 $k \in \{1, \dots, K\}$ 번째 쿼리 샘플을 나타낸다. M_k^* 는 실측 매칭 스코어를 나타낸다. 만일 p_k 가 실측 키포인트 페어이면, $M_k^* = 1$ 이고, 아니면 0이다.

[0076] 본 발명의 일 실시예에서는 $\phi(C, p)$ 를 직접 사용하여 예측된 키 포인트와 실제 키포인트 간에 엔드-포인트 에러를 사용한다. 보다 상세하게, 본 발명의 일 실시예에서는 채널별 평균 합산 비용 특징 볼륨 $V = \text{avgpool}(\phi(C, p)) \in \mathbb{R}^{H_s \times W_s \times H_t \times W_t}$ 을 도출하고, V에 soft avgmax 함수를 적용하여 F_{pred} 를 계산한다. 그리고, 실제 키포인트를 사용하여 계산된 실측 플로우 맵과 예측된 플로우 맵 사이의 유클리드 거리를 수학적 6과 같이 계산한다.

수학식 6

$$\mathcal{L}_c = \|F_{\text{gt}} - F_{\text{pred}}\|_2.$$

[0077]

수학식 5와 결합하여 밸런스된 가중치 λ_f 와 λ_c 를 가지는 최종 손실 함수 $\mathcal{L}_{\text{total}} = \lambda_f \mathcal{L}_f + \lambda_c \mathcal{L}_c$ 를 정의할 수 있다.

[0078]

도 4를 참조하여 추론 과정에 대해 설명하기로 한다.

[0079]

추론 과정에서 각 쿼리 $\mathbf{p} = [\mathbf{x}, \mathbf{y}]$ 를 위한 정확한 대응을 결정하기 위해 조밀한 대응 필드 $F(\mathbf{x})$ 를 찾는 것을 목표로 한다. 그러나 모든 대응 후보 $\mathbf{y}$ 에 대해 각 좌표 $\mathbf{x}$ 에 대한 최상의 대응을 검색하면 샘플당 피드포워드 $H_s \times W_s \times H_t \times W_t$ 수가 나오므로 시간이 많이 걸리고 계산 집약적이다.

[0080]

이 문제를 완화하기 위해 패치매치 기반 샘플링을 사용한다. 패치매치는 일련의 전파 및 업데이트 단계를 실행하여 검색 공간을 줄일 수 있다. 학습된 매칭 함수 f_θ 를 스코어 함수로 사용하여 대응을 결정할 수 있다.

[0081]

쿼리 $\mathbf{x}$ 에 대해 $(l-1)$ 번째 반복에서 인접 픽셀 $\mathbf{z}$ 에 대해 $\mathcal{Z}^{l-1} = \{F^{l-1}(\mathbf{z})\}_{\mathbf{z}}$ 가 되도록 인접 픽셀을 고려하여 후보 대응 세트를 샘플링한다.

[0082]

또한, 균일 분포에서 샘플링된 몇 개의 무작위 포인트는 $\mathcal{Y}^{l-1} = \bigcup (\mathcal{Z}^{l-1}, \{\mathbf{y}\})$ 이 되도록 $\mathcal{Z}^{l-1}$ 을 증가시키는 데 사용된다.

[0083]

대응 필드는 매칭 후보 $\mathcal{Y}^{l-1}$ 를 고려하여 수학식 7과 같이 도출될 수 있다.

[0084]

수학식 7

$$F^l(\mathbf{x}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{l-1}} (M([\mathbf{x}, \mathbf{y}]))$$

[0085]

해당 과정은 수렴될때까지 반복될 수 있다. 후보 선택 및 스코어링은 모든 대상 픽셀에 대해 병렬로 실행될 수 있다.

[0086]

또한, 본 발명의 일 실시예에 따르면, 연속 함수 f_θ 는 도 4에 도시된 바와 같이, 자연스럽게 미분 가능하기 때문에 해당 소스 좌표 $\mathbf{x}$ 에 대해 일치 점수의 음의 로그 구도를 줄이는 방향으로 타겟 좌표 $\mathbf{y}$ 를 최적화하기 위해 경사 하강법을 사용할 수 있다.

[0087]

좌표 최적화는 수학식 8과 같이 공식화될 수 있는 반복 업데이트에 의해 수행될 수 있다.

[0088]

수학식 8

$$\begin{aligned}\mathcal{L} &= -\log(M([\mathbf{x}, \mathbf{y}])), \\ \mathbf{y} &:= \mathbf{y} - \alpha \nabla_{\mathbf{y}} \mathcal{L},\end{aligned}$$

[0089]

[0090]

소스 좌표는 최적화 중에는 업데이트되지 않는다. 좌표 최적화를 통해 도 4와 같이 최종 대응 필드를 설정하기 위해 패치매치 기반 샘플링과 결합될 수 있으며, 반복 횟수는 N으로 설정될 수 있다.

[0091]

도 5에 예시된 바와 같이, 본 발명의 일 실시예에 따른 패치매치 기반 샘플링과 좌표 최적화에 의해 더 정교한 신경 매치 필드값을 예측할 수 있다.

[0093]

도 6은 본 발명의 일 실시예에 따른 컴퓨팅 장치의 내부 구성을 개략적으로 도시한 블록도이다.

[0094]

도 6을 참조하면, 본 발명의 일 실시예에 따른 컴퓨팅 장치(100)는 메모리(610) 및 프로세서(620)를 포함하여 구성된다.

[0095]

메모리(610)은 본 발명의 일 실시예에 따른 고정밀 대응을 위한 신경 매칭 표현 방법을 수행하기 위한 명령어들을 저장한다.

[0096]

프로세서(620)는 본 발명의 일 실시예에 따른 컴퓨팅 장치(100)의 내부 구성 요소들을 제어한다. 또한, 프로세서(620)는 메모리(610)에 저장된 명령어를 실행할 수 있다. 프로세서(620)에 의해 실행된 명령어들은 도 1을 참조하여 설명한 고정밀 대응을 위한 신경 매칭 표현 방법의 각각의 단계들을 수행할 수 있다. 이는 이미 기술한 바와 동일하므로 중복되는 설명은 생략하기로 한다.

[0098]

본 발명의 실시 예에 따른 장치 및 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 컴퓨터 판독 가능 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 분야 통상의 기술자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media) 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

[0099]

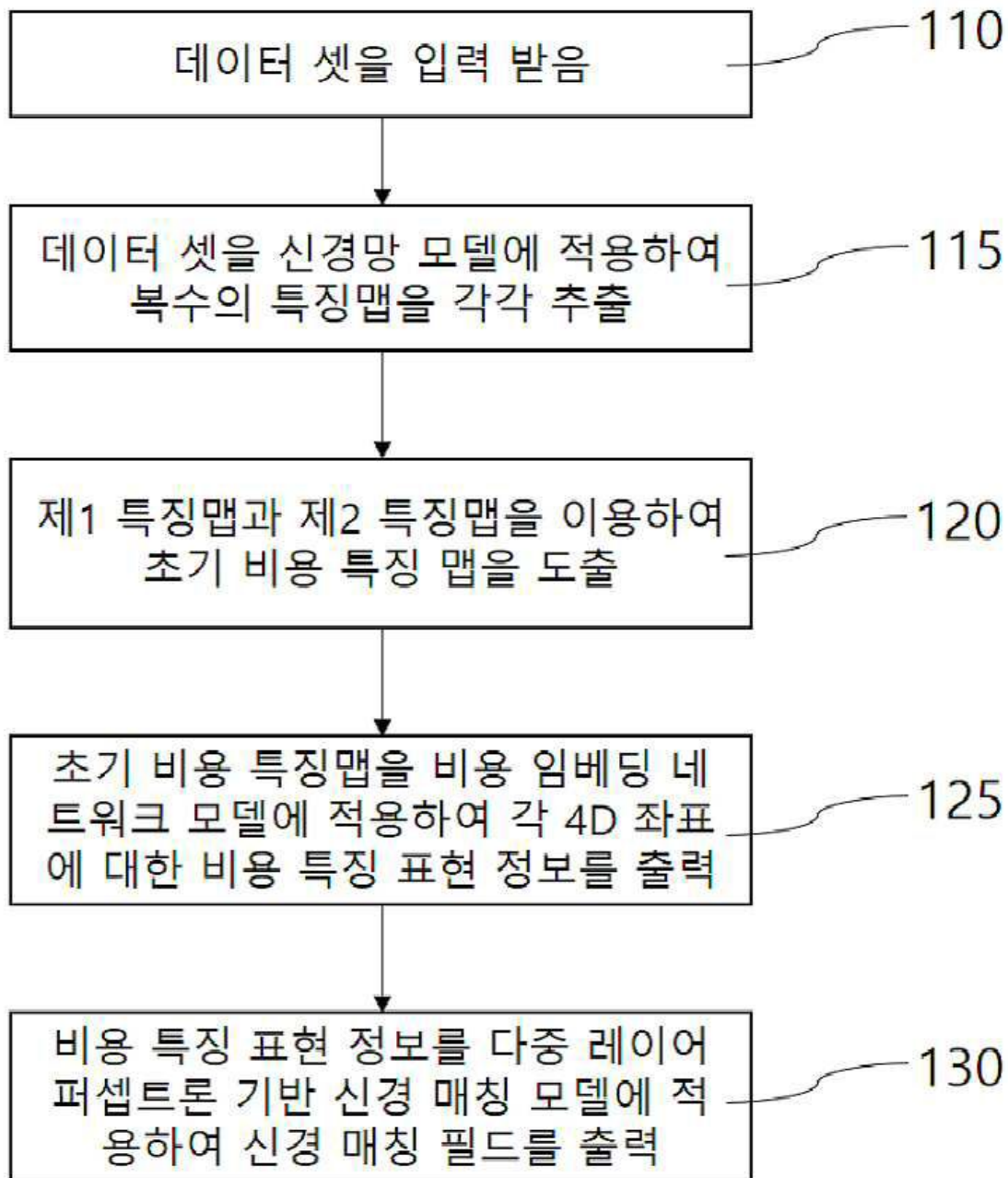
상술한 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0100]

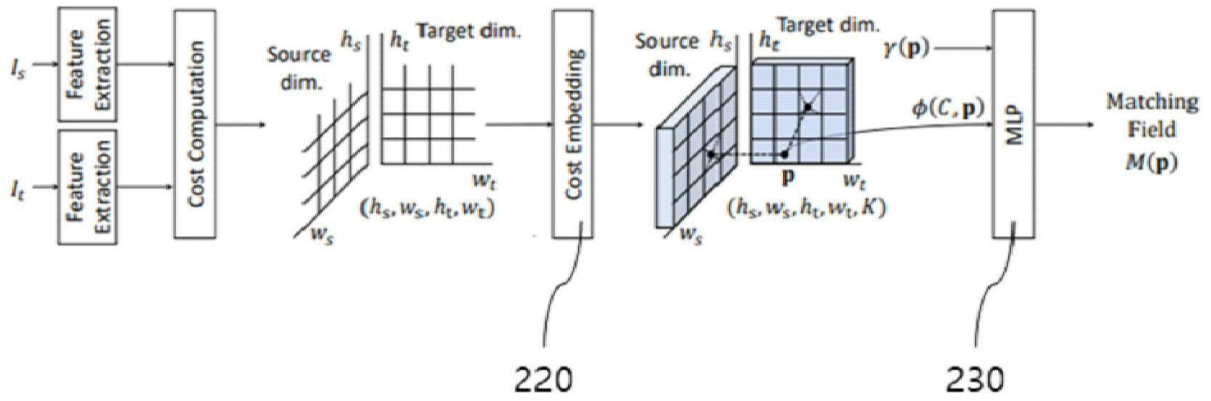
이제까지 본 발명에 대하여 그 실시 예들을 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

도면

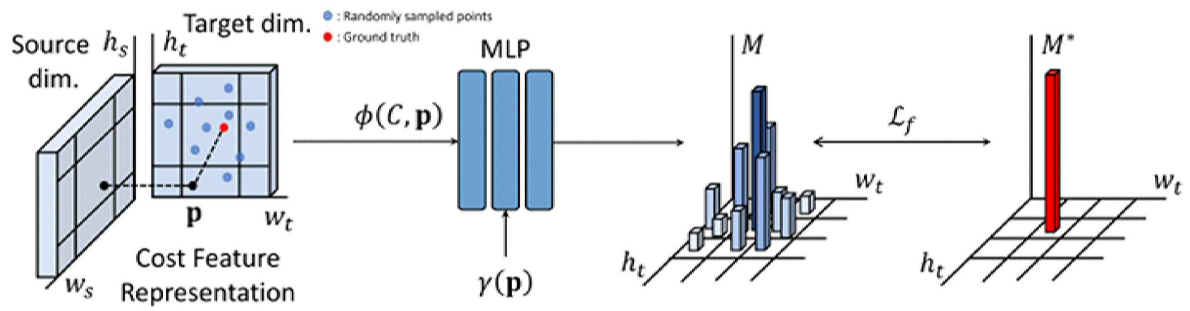
도면1



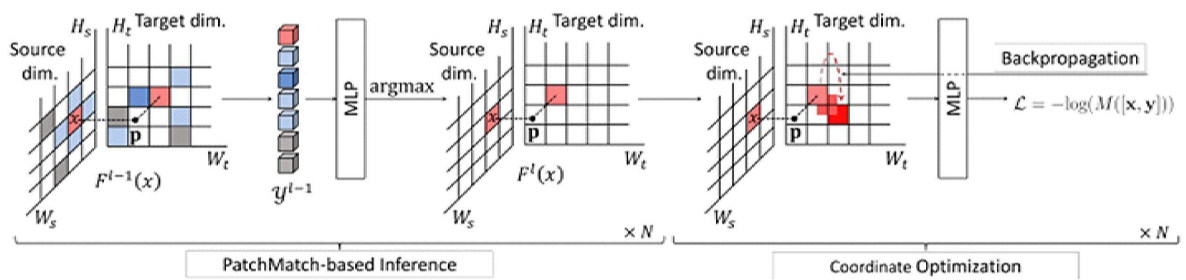
도면2



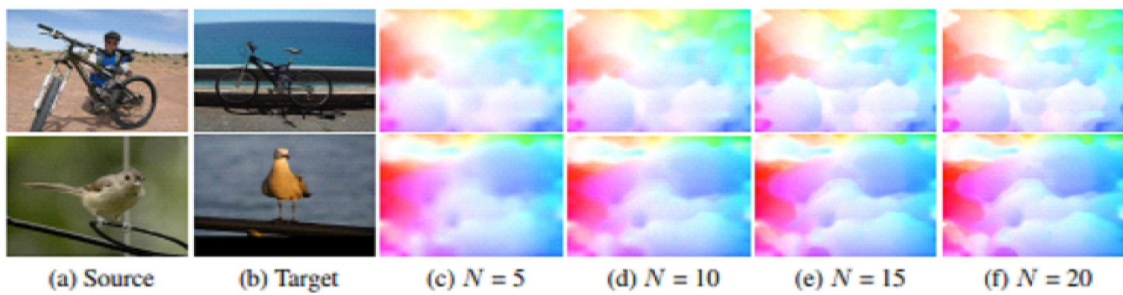
도면3



도면4



도면5



도면6

