(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0048266
(43) 공개일자 2025년04월08일

(51) 국제특허분류(Int. Cl.)
G06N 3/045 (2023.01) *G06N 3/091* (2023.01)
G06N 3/096 (2023.01) *G06V 10/77* (2022.01)
G06V 10/82 (2022.01)
(52) CPC특허분류
G06N 3/045 (2023.01)
G06N 3/091 (2023.01)
(21) 출원번호 10-2025-7005536
(22) 출원일자(국제) 2023년07월20일
심사청구일자 없음
(85) 번역문제출일자 2025년02월20일
(86) 국제출원번호 PCT/JP2023/026530
(87) 국제공개번호 WO 2024/029347
국제공개일자 2024년02월08일
(30) 우선권주장
JP-P-2022-125012 2022년08월04일 일본(JP)
JP-P-2022-170177 2022년10월24일 일본(JP)

(71) 출원인
소니 세미컨덕터 솔루션즈 가부시키가이샤
일본국 가나가와켄 아츠기시 아사히쵸 4-14-1
(72) 발명자
야마카와 겐지
일본 2430014 가나가와 아츠기시 아사히쵸 4-14-1 소니 세미컨덕터 솔루션즈 가부시키가이샤 시/오
가와사키 료헤이
일본 2430014 가나가와 아츠기시 아사히쵸 4-14-1 소니 세미컨덕터 솔루션즈 가부시키가이샤 시/오
(뒷면에 계속)
(74) 대리인
양영준, 김승식, 이중희

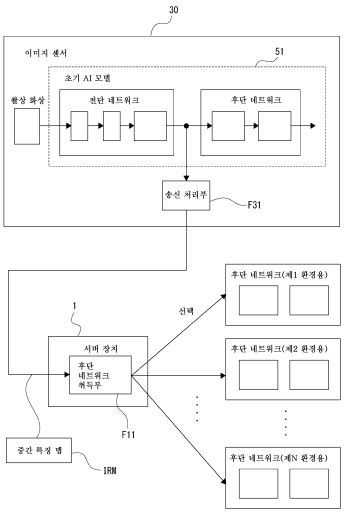
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **정보 처리 장치, 정보 처리 방법, 및 정보 처리 시스템**

(57) 요 약

본 기술에 관한 정보 처리 장치는, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 중간 특징 맵에 기초하여, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 후단 네트워크 취득부와, 후단 네트워크 취득부가 선택 또는 생성한 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 송신 처리부를 구비하고 있다.

대 표 도 - 도10



(52) CPC특허분류
G06N 3/096 (2023.01)
G06V 10/7715 (2023.08)
G06V 10/82 (2022.01)
(72) 발명자
에키 료지
일본 2430014 가나가와 아츠기시 아사히쵸 4-14-1
소니 세미컨덕터 솔루션즈 가부시키가이샤 시/오

야나기사와 에이타
일본 2430014 가나가와 아츠기시 아사히쵸 4-14-1
소니 세미컨덕터 솔루션즈 가부시키가이샤 시/오

명 세 서

청구범위

청구항 1

뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와,

상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 송신 처리부를 구비하는

정보 처리 장치.

청구항 2

제1항에 있어서,

상기 중간 특징 맵은, 상기 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터인, 정보 처리 장치.

청구항 3

제1항에 있어서,

상기 센서 장치는 촬상 센서이며, 상기 인공 지능 모델은 상기 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 된, 정보 처리 장치.

청구항 4

제3항에 있어서,

상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 된, 정보 처리 장치.

청구항 5

제3항에 있어서,

상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로 된, 정보 처리 장치.

청구항 6

제1항에 있어서,

상기 중간 특징 맵은, 상기 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 된, 정보 처리 장치.

청구항 7

제1항에 있어서,

상기 후단 네트워크 취득부는, 상기 중간 특징 맵에 기초하여, 상기 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하는, 정보 처리 장치.

청구항 8

제7항에 있어서,

상기 후단 네트워크의 복수의 후보는, 액티브 러닝으로서의 기계 학습에 의해 생성된, 정보 처리 장치.

청구항 9

제1항에 있어서,

상기 후단 네트워크 취득부는, 상기 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는, 정보 처리 장치.

청구항 10

제9항에 있어서,

상기 후단 네트워크 취득부는, 상기 중간 특징 맵에 기초하여, 상기 베이스가 되는 학습 완료 네트워크를 교사 모델로서 사용한 지식 증류를 행함으로써 하나의 상기 후단 네트워크를 생성하는, 정보 처리 장치.

청구항 11

정보 처리 장치가,

뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득 처리와,

상기 후단 네트워크 취득 처리에서 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 송신 처리를 행하는

정보 처리 방법.

청구항 12

뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 송신 처리부와,

상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 처리를 행하는 수신 처리부와,

상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부를 구비하는,

정보 처리 장치.

청구항 13

제12항에 있어서,

상기 중간 특징 맵은, 상기 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터인, 정보 처리 장치.

청구항 14

제12항에 있어서,

상기 센서 장치는 촬상 센서이며, 상기 인공 지능 모델은 상기 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데

이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 된, 정보 처리 장치.

청구항 15

제14항에 있어서,

상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 된, 정보 처리 장치.

청구항 16

제14항에 있어서,

상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로 된, 정보 처리 장치.

청구항 17

제12항에 있어서,

상기 중간 특징 맵은, 상기 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 된, 정보 처리 장치.

청구항 18

정보 처리 장치가,

뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 송신 처리와,

상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 수신 처리와,

상기 수신 처리에서 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리를 행하는

정보 처리 방법.

청구항 19

뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 제1 송신 처리부와,

상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와,

상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 제2 송신 처리부와,

상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 설정 데이터를 수신하는 처리를 행하는 수신 처리부와,

상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부 구비하는

정보 처리 시스템.

발명의 설명

기 술 분 야

[0001] 본 기술은, 정보 처리 장치와 그 방법, 및 정보 처리 시스템에 관한 것이며, 특히, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대하여, 센서 장치의 사용 환경 조건에 대한 적합화를 행하기 위한 기술에 관한 것이다.

배 경 기 술

[0002] 촬상 화상에 대하여 AI(Artificial Intelligence: 인공 지능) 모델을 사용한 처리인 AI 처리를 행하는 촬상 장치(이하 「AI 카메라」로 표기함)를 사용한 솔루션 비즈니스를 생각할 수 있다. 예를 들어, AI 카메라를 점포 내에 1개 또는 복수대 배치하고, 예를 들어 화상 인식 처리에 의해 고객의 속성(예를 들어 성별, 연령층 등)이나 행동 패턴에 관한 인식 결과를 얻고, 이 인식 결과를 인터넷 등을 통해 유저에게 제시한다고 하는 시스템을 제공하는 비즈니스 등을 생각할 수 있다.

[0003] AI 카메라측(에지측)에서 AI 처리를 행하기 때문에, 서버 장치측(클라우드측)이 각 카메라로부터 취득한 촬상 화상에 기초하여 AI 처리를 행하는 시스템과 비교하여, 처리 분산이 도모됨과 함께, 통신 데이터양의 저감을 도모할 수 있다.

[0004] 이때, AI 카메라가 갖는 AI 모델에 대하여, 서버 장치가 재학습을 행하는 서비스를 생각할 수 있다. 이 재학습 서비스에서는, AI 카메라가 놓인 환경, 예를 들어 일본 등의 고객의 인종이 한정되어 있는 국가의 점포에 배치되는지, 혹은 미국 등의 고객의 인종이 다양한 나라의 점포에 배치되는지 등과 같은 사용 지역의 차 등, AI 카메라의 사용 환경 조건에 따른 적합화를 위한 재학습을 행하는 것이 상정된다.

[0005] 또한, 관련되는 종래 기술에 대해서는 하기 특허문헌 1을 들 수 있다. 하기 특허문헌 1에는, DNN(Deep Neural Network)을 갖는 AI 모델을 제1 DNN 처리부와 제2 DNN 처리부로 분할하고, 전단의 제1 DNN 처리부에서 얻어지는 특징 맵을 입력 데이터로 하여, 후단의 제2 DNN 처리부가 추론 처리(예를 들어 물체 인식 처리)를 행하는 구성이 개시되어 있다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 일본 특허 공개 제2020-47191호 공보

발명의 내용

해결하려는 과제

[0007] 여기서, 상기한 재학습 서비스에서는, AI 모델 전체를 재학습한다고 하는 방법을 채용하는 것을 생각할 수 있다.

[0008] 그러나, AI 모델 전체를 재학습하기 위해서는 필요한 학습 횟수가 증대되는 경향이 되고, 그 결과, AI 카메라(에지측)에 있어서의 AI 모델을 AI 카메라의 사용 환경 조건에 적합시키기 위해 많은 시간을 요하는 것이 되어 버린다. 또한, 필요한 학습 횟수가 증대된다고 하는 것은, 준비해야 할 학습용 입력 데이터의 수도 증대화되게 된다. 재학습을 위한 학습용 입력 데이터를 재학습 대상의 AI 모델의 이용자가 클라우드측에 업로드하는 것을 전제로 한 경우, 학습용 입력 데이터수가 증대된다고 하는 것은, 이용자가 클라우드측에 송신해야 할 데이터양의 증대화를 초래하는 것이 되고, 결과, AI 모델을 AI 카메라의 사용 환경 조건에 적합화시키기 위해 요하는 통신 데이터양의 증대화를 초래해 버린다.

[0009] 또한, AI 모델을 AI 카메라의 사용 환경 조건에 적합화시키기 위한 재학습을 실현하는 데 있어서는, AI 카메라에 의한 실제의 촬상 화상을 클라우드측에 송신한다고 하는 것도 생각할 수 있지만, 촬상 화상 그 자체를 송신하는 것은 개인 정보의 유출을 초래할 우려가 있어 바람직하지 않다.

[0010] 본 기술은 상기 사정을 감안하여 이루어진 것이며, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에

적합화함에 있어서의 소요 시간 단축화, 및 통신 데이터양의 삭감, 및 개인 정보의 유출 가능성의 저감을 도모하는 것을 목적으로 한다.

과제의 해결 수단

[0011] 본 기술에 관한 제1 정보 처리 장치는, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와, 상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 송신 처리부를 구비하는 것이다.

[0012] 상기 구성에 의하면, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서, 인공 지능 모델의 네트워크 전체가 아니라 후단 네트워크에 대해서만 선택의 후보를 작성해 두면 된다. 또는, 베이스가 되는 학습 완료 네트워크로부터 예를 들어 증류 등에 의해 대응하는 후단 네트워크를 생성하면 된다. 이때, 후보가 되는 후단 네트워크의 생성에 요하는 학습 횟수는, 전단 네트워크를 포함하는 네트워크 전체의 재학습에 요하는 학습 횟수보다도 적게 하는 것이 가능하다. 또한, 베이스가 되는 학습 완료 네트워크로부터 대응하는 후단 네트워크를 생성하는 데 요하는 학습 횟수에 대해서도, 네트워크 전체의 재학습에 요하는 학습 횟수보다도 적게 하는 것이 가능하다. 필요한 학습 횟수를 적게 할 수 있음으로써, 센서 장치의 사용 환경 조건에 적합한 후단 네트워크의 생성에 요하는 시간의 단축화를 도모할 수 있다. 또한, 필요한 학습 횟수를 적게 할 수 있음으로써, 센서 장치의 사용 환경 조건에 적합한 후단 네트워크의 생성에 요하는 학습용 입력 데이터수의 삭감을 도모할 수 있다.

[0013] 또한, 상기 구성에 의하면, 센서 장치의 사용 환경 조건에 적합한 하나의 후단 네트워크를 취득함에 있어서는, 에지측으로부터 중간 특징 맵을 취득하면 된다. 중간 특징 맵은 입력 데이터가 중간층에서 처리되어 얻어진 데이터이기 때문에, 그대로는 개인 정보를 특정하는 것이 곤란한 데이터가 된다.

[0014] 또한, 본 기술에 관한 제1 정보 처리 방법은, 정보 처리 장치가, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득 처리와, 상기 후단 네트워크 취득 처리에서 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 송신 처리를 행하는 정보 처리 방법이다.

[0015] 이와 같은 제1 정보 처리 방법에 의해서도, 상기한 본 기술에 관한 제1 정보 처리 장치와 마찬가지의 작용이 얻어진다.

[0016] 본 기술에 관한 제2 정보 처리 장치는, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 송신 처리부와, 상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 처리를 행하는 수신 처리부와, 상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부를 구비하는 것이다.

[0017] 상기 구성에 의하면, 센서 장치의 검출 데이터에 기초하는 추론 처리를 행하는 에지측의 정보 처리 장치로서, 외부 장치가 센서 장치의 사용 환경 조건에 적합한 후단 네트워크를 구하기 위한 방법에, 중간 특징 맵에 기초하여 복수의 후단 네트워크의 후보 중에서 하나의 후단 네트워크를 선택하는 방법, 또는, 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 방법 중 어느 것을 채용하는 것을 가능하게 하는 정보 처리 장치가 제공된다.

[0018] 또한, 본 기술에 관한 제2 정보 처리 방법은, 정보 처리 장치가, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출

데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 송신 처리와, 상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 수신 처리와, 상기 수신 처리에서 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리를 행하는 정보 처리 방법이다.

[0019] 이와 같은 제2 정보 처리 방법에 의해서도, 상기한 본 기술에 관한 제2 정보 처리 장치와 마찬가지의 작용이 얻어진다.

[0020] 본 기술에 관한 정보 처리 시스템은, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 제1 송신 처리부와, 상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와, 상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 제2 송신 처리부와, 상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 설정 데이터를 수신하는 처리를 행하는 수신 처리부와, 상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부를 구비하는 것이다.

[0021] 이와 같은 정보 처리 시스템에 의해서도, 상기한 본 기술에 관한 제1, 제2 정보 처리 장치와 마찬가지의 작용이 얻어진다.

도면의 간단한 설명

[0022] 도 1은 실시 형태로서의 정보 처리 시스템의 개요 구성예를 도시한 블록도이다.

도 2는 실시 형태로서의 정보 처리 시스템이 구비하는 정보 처리 장치의 하드웨어 구성예를 도시한 블록도이다.

도 3은 실시 형태로서의 센서 장치를 구비한 카메라의 구성예를 도시한 블록도이다.

도 4는 실시 형태로서의 센서 장치의 구조예에 대한 설명도이다.

도 5는 실시 형태에 있어서의 AI 모델의 개요의 설명도이다.

도 6은 실시 형태에 있어서의 초기 AI 모델의 설명도이다.

도 7은 초기 AI 모델의 분단, 및 전단 네트워크와 후단 네트워크의 설명도이다.

도 8은 후단 네트워크의 후보의 생성 방법예에 대한 설명도이다.

도 9는 재학습에 의해 생성된 후단 네트워크의 복수의 후보에 대한 설명도이다.

도 10은 후단 네트워크의 선택에 관한 서버 장치, 이미지 센서 각각의 기능을 설명하기 위한 도면이다.

도 11은 중간 특징 맵에 기초하여 하나의 후단 네트워크를 선택하는 방법의 일례에 대한 설명도이다.

도 12는 선택한 후단 네트워크의 디플로이에 관한 서버 장치, 이미지 센서 각각이 갖는 기능의 설명도이다.

도 13은 초기 AI 모델에 대하여 2개의 다른 분단 위치에서 얻어지는 중간 특징 맵의 가시화 이미지를 예시한 도면이다.

도 14는 중간 특징 맵에 있어서의 개인 정보에 관한 화상 영역에 대한 설명도이다.

도 15는 분단 위치의 다른 설정 방법에 대한 설명도이다.

도 16은 실시 형태로서의 적합화 방법을 실현하기 위한 처리 수순예를 나타낸 흐름도이다.

도 17은 액티브 러닝을 적용한 후단 네트워크의 생성 방법에 대한 설명도이다.

도 18은 지식 증류에 의한 소형 후단 네트워크의 생성 방법의 일례에 대한 설명도이다.

도 19는 지식 증류에 의한 소형 후단 네트워크의 생성 방법의 다른 예에 대한 설명도이다.

도 20은 지식 증류에 의한 소형 후단 네트워크의 생성을 행하는 경우에 대응한 적합화 방법 실현을 위한 처리 수순예를 나타낸 흐름도이다.

도 21은 실시 형태로서의 정보 처리 시스템의 운용예에 대하여 설명하기 위한 도면이다.

도 22는 실시 형태로서의 정보 처리 시스템의 다른 운용예에 대하여 설명하기 위한 도면이다.

도 23은 AI 커스터마이즈의 변형예에 대한 설명도이다.

도 24는 전단 네트워크의 출력을 촬상 설정의 제어에 피드백하는 변형예의 설명도이다.

발명을 실시하기 위한 구체적인 내용

[0023] 이하, 첨부 도면을 참조하여, 본 기술에 관한 실시 형태를 다음의 순서로 설명한다.

[0024] <1. 정보 처리 시스템의 개요>

[0025] <2. 정보 처리 장치의 구성>

[0026] <3. 실시 형태로서의 적합화 방법>

[0027] <4. 처리 수순>

[0028] <5. 후단 네트워크의 소형화에 대하여>

[0029] <6. 시스템 운용예>

[0030] <7. AI 커스터마이즈의 변형예>

[0031] <8. 변형예>

[0032] <9. 실시 형태의 정리>

[0033] <10. 본 기술>

[0034] <1. 정보 처리 시스템의 개요>

[0035] 도 1은 본 기술에 관한 실시 형태로서의 정보 처리 시스템(100)의 개략 구성예를 도시한 블록도이다.

[0036] 도시와 같이 정보 처리 시스템(100)은, 서버 장치(1)와, 1개 또는 복수의 유저 단말기(2)와, 복수의 카메라(3)와, 포그 서버(4)를 구비하고 있다. 본 예에 있어서, 서버 장치(1)는, 유저 단말기(2) 및 포그 서버(4)와의 사이에서 예를 들어 인터넷 등으로 된 네트워크(5)를 통한 상호 통신을 행하는 것이 가능하게 구성되어 있다.

[0037] 서버 장치(1), 유저 단말기(2), 및 포그 서버(4)는, CPU(Central Processing Unit), ROM(Read Only Memory), 및 RAM(Random Access Memory)을 갖는 마이크로컴퓨터를 구비한 정보 처리 장치로서 구성되어 있다.

[0038] 여기서, 유저 단말기(2)는, 정보 처리 시스템(100)을 사용한 서비스의 수신자인 유저에 의해 사용되는 것이 상정되는 정보 처리 장치이다. 또한, 서버 장치(1)는, 서비스의 제공자에 의해 사용되는 것이 상정되는 정보 처리 장치이다.

[0039] 각 카메라(3)는, 예를 들어 CCD(Charge Coupled Device)형 이미지 센서나 CMOS(Complementary Metal Oxide Semiconductor)형 이미지 센서 등의 이미지 센서를 구비하고, 피사체를 촬상하여 디지털 데이터로서의 화상 데이터(촬상 화상 데이터)를 얻는다. 또한, 후술하는 바와 같이 각 카메라(3)는, 촬상 화상에 대하여 AI(Artificial Intelligence: 인공 지능) 모델을 사용한 처리인 AI 처리를 행하는 기능도 갖고 있다.

[0040] 각 카메라(3)는, 포그 서버(4)와 데이터 통신 가능하게 구성되며, 예를 들어 AI 모델을 사용한 화상 처리의 결과를 나타내는 처리 결과 정보 등의 각종 데이터를 포그 서버(4)에 송신하거나, 포그 서버(4)로부터 각종 데이터를 수신하거나 하는 것이 가능하게 된다.

[0041] 여기서, 도 1에 도시한 정보 처리 시스템(100)에 대해서는, 각 카메라(3)의 AI 처리에서 얻어지는 AI 처리 결과

를 나타내는 정보(이하, 「처리 결과 정보」로 표기함)에 기초하여, 포그 서버(4) 또는 서버 장치(1)가 피사체의 분석 정보를 생성하고, 생성한 분석 정보를 유저 단말기(2)를 통해 유저에게 열람시킨다고 하는 용도가 상정된다.

[0042] 이 경우, 각 카메라(3)의 용도로서는, 각종 감시 카메라의 용도를 생각할 수 있다. 예를 들어, 점포나 오피스, 주택 등의 옥내에 대한 감시 카메라, 주차장이나 거리 등의 옥외를 감시하기 위한 감시 카메라(교통 감시 카메라 등을 포함함), FA(Factory Automation)나 IA(Industrial Automation)에 있어서의 제조 라인의 감시 카메라, 차내나 차외를 감시하는 감시 카메라 등의 용도를 들 수 있다.

[0043] 예를 들어, 점포에 있어서의 감시 카메라의 용도이면, 복수의 카메라(3)를 점포 내의 소정 위치에 각각 배치하고, 유저가 방문객의 고객층(성별이나 연령층 등)이나 점포 내에서의 행동(동선) 등을 확인할 수 있도록 하는 것을 생각할 수 있다. 그 경우, 상기한 분석 정보로서는, 이들 방문객의 고객층의 정보나 점포 내에서의 동선의 정보, 및 정산 계산대에 있어서의 혼잡 상태의 정보(예를 들어, 정산 계산대의 대기 시간 정보) 등을 생성하는 것을 생각할 수 있다.

[0044] 혹은, 교통 감시 카메라의 용도이면, 각 카메라(3)를 도로 근방의 각 위치에 배치하고, 유저가 통과 차량에 대한 넘버(차량 번호)나 차의 색, 차종 등의 정보를 인식할 수 있도록 하는 것을 생각할 수 있고, 그 경우, 상기한 분석 정보로서는, 이들 넘버나 차의 색, 차종 등의 정보를 생성하는 것을 생각할 수 있다.

[0045] 또한, 주차장에 교통 감시 카메라를 사용한 경우에는, 주차되어 있는 각 차량을 감시할 수 있도록 카메라를 배치하고, 의심스러운 행동을 하고 있는 수상한 사람이 각 차량의 주변에 없는지 등을 감시하고, 수상한 사람이 있는 경우에는, 수상한 사람이 있는 것이나 그 수상한 사람의 속성(성별이나 연령층, 복장 등) 등을 통지하는 것을 생각할 수 있다.

[0046] 또한, 거리나 주차장의 빈 스페이스를 감시하여, 유저에게 차를 주차할 수 있는 스페이스의 장소를 통지하는 것 등도 생각할 수 있다.

[0047] 포그 서버(4)는, 예를 들어 상기한 점포의 감시 용도에 있어서는 각 카메라(3)와 함께 감시 대상의 점포 내에 배치되는 등, 감시 대상마다 배치되는 것이 상정된다. 이와 같이 점포 등의 감시 대상마다 포그 서버(4)를 마련함으로써, 감시 대상에 있어서의 복수의 카메라(3)로부터의 송신 데이터를 서버 장치(1)가 직접 수신할 필요가 없어져, 서버 장치(1)의 처리 부담 경감이 도모된다.

[0048] 또한, 포그 서버(4)는, 감시 대상으로 하는 점포가 복수 있고, 그것들 점포가 모두 동일 계열에 속하는 점포인 경우에는, 점포마다 마련하는 것이 아니라, 그것들 복수의 점포에 대하여 하나 마련하는 것도 생각할 수 있다. 즉, 포그 서버(4)는, 감시 대상마다 하나 마련하는 것에 한정되지는 않고, 복수의 감시 대상에 대하여 하나의 포그 서버(4)를 마련하는 것도 가능한 것이다.

[0049] 또한, 정보 처리 시스템(100)으로서는, 포그 서버(4)를 생략하고, 각 카메라(3)를 직접 네트워크(5)에 접속시켜, 복수의 카메라(3)로부터의 송신 데이터를 서버 장치(1)가 직접 수신하는 구성도 채용할 수 있다.

[0050] 서버 장치(1)는, 정보 처리 시스템(100)을 통괄적으로 관리하는 기능을 갖는 정보 처리 장치로 된다.

[0051] 서버 장치(1)는, 정보 처리 시스템(100)의 관리에 관한 기능으로서, 도시된 바와 같이 라이선스 오서리제이션 기능 F1, 어카운트 서비스 기능 F2, 및 AI 서비스 기능 F3을 갖는다.

[0052] 라이선스 오서리제이션 기능 F1은, 각종 인증에 관한 처리를 행하는 기능이다. 구체적으로, 라이선스 오서리제이션 기능 F1에서는, 각 카메라(3)의 디바이스 인증에 관한 처리나, 카메라(3)에서 사용되는 AI 모델 등의 데이터에 대한 인증에 관한 처리가 행해진다.

[0053] 라이선스 오서리제이션 기능 F1에 있어서, 카메라(3)의 인증에 대해서는, 카메라(3)와 네트워크(5)를 통해 접속된 경우(본 예에서는 포그 서버(4)를 통한 접속으로 됨)에, 카메라(3)마다 디바이스 ID를 발행하는 처리가 행해진다.

[0054] 또한, AI 모델의 인증에 대해서는, 등록 신청된 AI 모델에 대하여, 고유의 ID(AI 모델 ID)를 발행하는 처리가 행해진다.

[0055] 또한, 라이선스 오서리제이션 기능 F1에서는, 카메라(3)와 서버 장치(1) 사이에서 시큐어한 통신이 행해지도록 하기 위한 각종 키나 증명서 등을 카메라(3)의 제조업자(특히 후술하는 이미지 센서(30)의 제조업자)에게 발행하는 처리가 행해짐과 함께, 증명 효력의 정지나 갱신을 위한 처리도 행해진다.

[0056] 또한, 라이선스 오서리제이션 기능 F1에서는, 이하에서 설명하는 어카운트 서비스 기능 F2에 의해 유저 등록(유저 ID의 발행을 수반하는 어카운트 정보의 등록)이 행해진 경우에, 유저가 구입한 카메라(3)(상기 디바이스 ID)와 유저 ID를 관련짓는 처리도 행해진다.

[0057] 어카운트 서비스 기능 F2는, 유저의 어카운트 정보의 생성이나 관리를 행하는 기능이다. 어카운트 서비스 기능 F2에서는, 유저 정보의 입력을 접수하고, 입력된 유저 정보에 기초하여 어카운트 정보를 생성한다(적어도 유저 ID와 패스워드 정보를 포함하는 어카운트 정보의 생성을 행한다).

[0058] AI 서비스 기능 F3은, AI 카메라로서의 카메라(3)의 이용에 관한 서비스를 유저에게 제공하기 위한 기능으로 된다.

[0059] 이 AI 서비스 기능 F3의 하나로서는, 유저로부터의 지시에 기초하여, 카메라(3)에 대하여 AI 모델을 디플로이하는 기능을 들 수 있다. 여기에서 말하는 디플로이란, 대상 장치에 AI 모델을 사용 가능하게 인스톨시키기 위한 송신 처리를 의미한다.

[0060] 또한, AI 서비스 기능 F3의 하나로는, 전술한 분석 정보의 생성에 관한 기능을 들 수도 있다. 즉, 카메라(3)에 있어서의 AI 처리의 처리 결과 정보에 기초하여 피사체의 분석 정보를 생성하고, 생성한 분석 정보를 유저 단말기(2)를 통해 유저에게 열람시키기 위한 처리를 행하는 기능이다.

[0061] 또한, AI 서비스 기능 F3의 하나로는, AI 모델의 재학습 기능을 들 수 있다. 즉, 카메라(3)에 인스톨된 AI 모델에 대한 재학습 기능이다. 본 실시 형태에서는, 이 재학습 기능에 의해, AI 모델을 카메라(3)(이미지 센서)의 사용 환경 조건에 적합시키기 위한 처리가 행해지지만, 이 점에 대해서는 후에 다시 설명한다.

[0062] 여기서, 상기에서는, 서버 장치(1) 단체로 라이선스 오서리제이션 기능 F1, 어카운트 서비스 기능 F2, 및 AI 서비스 기능 F3을 실현하는 구성을 예시하였지만, 이들 기능을 복수의 정보 처리 장치가 분담하여 실현하는 구성으로 하는 것도 가능하다. 예를 들어, 상기 기능을 각각 1대의 정보 처리 장치가 담당하는 구성으로 하는 것을 생각할 수 있다. 혹은, 상기한 기능 중 하나의 기능을 복수의 정보 처리 장치가 분담한다고 하는 구성을 취하는 것도 가능하다.

[0063] <2. 정보 처리 장치의 구성>

[0064] 도 2는 서버 장치(1)의 하드웨어 구성예를 도시한 블록도이다.

[0065] 또한, 유저 단말기(2) 및 포그 서버의 하드웨어 구성예도 이 도 2에 도시한 것과 마찬가지로 된다.

[0066] 도시와 같이 서버 장치(1)는, CPU(11)를 구비하고 있다. CPU(11)는, 지금까지 서버 장치(1)의 처리로서 설명한 각종 처리를 행하는 연산 처리부로서 기능하고, ROM(12)이나 예를 들어 EEP-ROM(Electrically Erasable Programmable Read-Only Memory) 등의 불휘발성 메모리부(14)에 기억되어 있는 프로그램, 또는 기억부(19)로부터 RAM(13)에 로드된 프로그램에 따라서 각종 처리를 실행한다. RAM(13)에는 또한, CPU(11)가 각종 처리를 실행하는 데 있어서 필요한 데이터 등도 적절히 기억된다.

[0067] CPU(11), ROM(12), RAM(13), 및 불휘발성 메모리부(14)는, 버스(23)를 통해 서로 접속되어 있다. 이 버스(23)에는 또한, 입출력 인터페이스(I/F)(15)도 접속되어 있다.

[0068] 입출력 인터페이스(15)에는, 조작자나 조작 디바이스로 이루어지는 입력부(16)가 접속된다. 예를 들어, 입력부(16)로서는, 키보드, 마우스, 키, 다이얼, 터치 패널, 터치 패드, 리모트 컨트롤러 등의 각종 조작자나 조작 디바이스가 상정된다.

[0069] 입력부(16)에 의해 유저의 조작이 검지되고, 입력된 조작에 따른 신호는 CPU(11)에 의해 해석된다.

[0070] 또한 입출력 인터페이스(15)에는, LCD(Liquid Crystal Display) 혹은 유기 EL(Electro-Luminescence) 디스플레이 등으로 이루어지는 표시부(17)나, 스피커 등으로 이루어지는 음성 출력부(18)가 일체 또는 별체로서 접속된다.

[0071] 표시부(17)는 각종 정보 표시에 사용되며, 예를 들어 컴퓨터 장치의 하우징에 마련되는 디스플레이 디바이스나, 컴퓨터 장치에 접속되는 별체의 디스플레이 디바이스 등에 의해 구성된다.

[0072] 표시부(17)는, CPU(11)의 지시에 기초하여 표시 화면 상에 각종 화상 처리를 위한 화상이나 처리 대상의 동화상 등의 표시를 실행한다. 또한 표시부(17)는 CPU(11)의 지시에 기초하여, 각종 조작 메뉴, 아이콘, 메시지 등, 즉 GUI(Graphical User Interface)로서의 표시를 행한다.

[0073] 입출력 인터페이스(15)에는, HDD(Hard Disk Drive)나 고체 메모리 등으로 구성되는 기억부(19)나, 모뎀 등으로 구성되는 통신부(20)가 접속되는 경우도 있다.

[0074] 통신부(20)는, 인터넷 등의 전송로를 통한 통신 처리나, 각종 기기와의 유선/무선 통신, 버스 통신 등에 의한 통신을 행한다.

[0075] 입출력 인터페이스(15)에는 또한, 필요에 따라서 드라이브(21)가 접속되고, 자기 디스크, 광 디스크, 광자기 디스크, 혹은 반도체 메모리 등의 리무버블 기록 매체(22)가 적절히 장착된다.

[0076] 드라이브(21)에 의해, 리무버블 기록 매체(22)로부터 각 처리에 사용되는 프로그램 등의 데이터 파일 등을 읽어낼 수 있다. 읽어내진 데이터 파일은 기억부(19)에 기억되거나, 데이터 파일에 포함되는 화상이나 음성이 표시부(17)나 음성 출력부(18)에서 출력되거나 한다. 또한 리무버블 기록 매체(22)로부터 읽어내진 컴퓨터 프로그램 등은 필요에 따라서 기억부(19)에 인스톨된다.

[0077] 상기와 같은 하드웨어 구성을 갖는 컴퓨터 장치에서는, 예를 들어 본 실시 형태의 처리를 위한 소프트웨어를, 통신부(20)에 의한 네트워크 통신이나 리무버블 기록 매체(22)를 통해 인스톨할 수 있다. 혹은, 당해 소프트웨어는 미리 ROM(12)이나 기억부(19) 등에 기억되어 있어도 된다.

[0078] CPU(11)가 각종 프로그램에 기초하여 처리 동작을 행함으로써, 서버 장치(1)로서의 필요한 정보 처리나 통신 처리가 실행된다.

[0079] 또한, 서버 장치(1)는, 도 2와 같은 컴퓨터 장치가 단일로 구성되는 것에 한하지 않고, 복수의 컴퓨터 장치가 시스템화되어 구성되어도 된다. 복수의 컴퓨터 장치는, LAN(Local Area Network) 등에 의해 시스템화되어 있어도 되고, 인터넷 등을 이용한 VPN(Virtual Private Network) 등에 의해 원격지에 배치된 것이어도 된다. 복수의 컴퓨터 장치에는, 클라우드 컴퓨팅 서비스에 의해 이용 가능한 서버 군(클라우드)으로서의 컴퓨터 장치가 포함되어도 된다.

[0080] 도 3은 카메라(3)의 구성예를 도시한 블록도이다.

[0081] 도시와 같이 카메라(3)는, 이미지 센서(30), 촬상 광학계(31), 광학계 구동부(32), 제어부(33), 메모리부(34), 통신부(35), 및 센서부(36)를 구비하고 있다. 이미지 센서(30), 제어부(33), 메모리부(34), 통신부(35), 및 센서부(36)는, 버스(37)를 통해 접속되어, 서로 데이터 통신을 행하는 것이 가능하게 되어 있다.

[0082] 촬상 광학계(31)는, 커버 렌즈, 줌 렌즈, 포커스 렌즈 등의 렌즈나 조리개(아이리스) 기구를 구비한다. 이 촬상 광학계(31)에 의해, 피사체로부터의 광(입사광)이 유도되어, 이미지 센서(30)의 수광면에 집광된다.

[0083] 광학계 구동부(32)는, 촬상 광학계(31)가 갖는 줌 렌즈, 포커스 렌즈, 및 조리개 기구의 구동부를 포괄적으로 나타낸 것이다. 구체적으로, 광학계 구동부(32)는, 이들 줌렌즈, 포커스 렌즈, 조리개 기구 각각을 구동하기 위한 액추에이터, 및 해당 액추에이터의 구동 회로를 갖고 있다.

[0084] 제어부(33)는, 예를 들어 CPU, ROM, 및 RAM을 갖는 마이크로컴퓨터를 구비하여 구성되며, CPU가 ROM에 기억되어 있는 프로그램, 또는 RAM에 로드된 프로그램에 따라서 각종 처리를 실행함으로써, 카메라(3)의 전체 제어를 행한다.

[0085] 또한, 제어부(33)는, 광학계 구동부(32)에 대하여 줌 렌즈, 포커스 렌즈, 조리개 기구 등의 구동 지시를 행한다. 광학계 구동부(32)는 이들 구동 지시에 따라서 포커스 렌즈나 줌 렌즈의 이동, 조리개 기구의 조리개 블레이드의 개폐 등을 실행시키게 된다.

[0086] 또한, 제어부(33)는, 메모리부(34)에 대한 각종 데이터의 기입이나 읽어내기에 대한 제어를 행한다.

[0087] 메모리부(34)는, 예를 들어 HDD나 플래시 메모리 장치 등의 불휘발성의 기억 디바이스로 되어, 제어부(33)가 각종 처리를 실행함에 있어서 사용하는 데이터의 기억에 사용된다. 또한, 메모리부(34)는, 이미지 센서(30)로부터 출력된 화상 데이터의 보존처(기록처)로서도 사용할 수 있다.

[0088] 제어부(33)는, 통신부(35)를 통해 외부 장치와의 사이에서 각종 데이터 통신을 행한다. 본 예에 있어서의 통신부(35)는, 적어도 도 1에 도시한 포그 서버(4)와의 사이에서의 데이터 통신을 행하는 것이 가능하게 구성되어 있다.

[0089] 혹은, 통신부(35)로서는, 네트워크(5)를 통한 통신이 가능하게 되어, 서버 장치(1)와의 사이에서 데이터 통신이 행해지는 경우도 있다.

[0090] 센서부(36)는, 카메라(3)가 구비하는 이미지 센서(30) 이외의 센서를 포괄적으로 나타내고 있다. 센서부(36)가 갖는 센서로서는, 예를 들어 카메라(3)의 위치나 고도를 검출하기 위한 GNSS(Global Navigation Satellite System) 센서나 고도 센서, 환경 온도를 검출하기 위한 온도 센서, 카메라(3)의 움직임을 검출하기 위한 가속도 센서나 각속도 센서 등의 움직임 센서 등을 들 수 있다.

[0091] 이미지 센서(30)는, 예를 들어 CCD형, CMOS형 등의 고체 촬상 소자로서 구성되며, 도시와 같이 촬상부(41), 화상 신호 처리부(42), 센서 내 제어부(43), AI 처리부(44), 메모리부(45), 컴퓨터 비전 처리부(46), 및 통신 인터페이스(I/F)(47)를 구비하고 있다. 이들 각 부는 각각이 버스(48)를 통해 서로 데이터 통신 가능하게 되어 있다.

[0092] 이 이미지 센서(30)는, 본 기술에 관한 정보 처리 장치의 일 실시 형태이다.

[0093] 촬상부(41)는, 포토다이오드 등의 광전 변환 소자를 갖는 화소가 2차원으로 배열된 화소 어레이부와, 화소 어레이부가 구비하는 각 화소로부터 광전 변환에 의해 얻어진 전기 신호를 읽어내는 읽어내기 회로를 구비하고 있다. 읽어내기 회로에서는, 광전 변환에 의해 얻어진 전기 신호에 대하여, 예를 들어 CDS(Correlated Double Sampling) 처리, AGC(Automatic Gain Control) 처리 등을 실행하고, 또한 A/D(Analog to Digital) 변환 처리를 행한다.

[0094] 화상 신호 처리부(42)는, A/D 변환 처리 후의 디지털 데이터로서의 촬상 화상 신호에 대하여, 전처리, 동시화 처리, YC 생성 처리, 해상도 변환 처리, 코덱 처리 등을 행한다.

[0095] 전처리에서는, 촬상 화상 신호에 대하여 R(적), G(녹), B(청)의 흑 레벨을 소정의 레벨로 클램프하는 클램프 처리나, R, G, B의 색 채널간의 보정 처리 등을 행한다. 동시화 처리에서는, 각 화소에 대한 화상 데이터가, R, G, B 모든 색 성분을 갖도록 하는 색 분리 처리를 실시한다. 예를 들어, 베이어 배열의 컬러 필터를 사용한 촬상 소자의 경우에는, 색 분리 처리로서 디모자이크 처리가 행해진다. YC 생성 처리에서는, R, G, B의 화상 데이터로부터, 휘도(Y) 신호 및 색(C) 신호를 생성(분리)한다. 해상도 변환 처리에서는, 각종 신호 처리가 실시된 화상 데이터에 대하여, 해상도 변환 처리를 실행한다.

[0096] 코덱 처리에서는, 상기 각종 처리가 실시된 화상 데이터에 대하여, 예를 들어 기록용이나 통신용의 부호화 처리, 파일 생성을 행한다. 코덱 처리에서는, 동화상의 파일 형식으로서, 예를 들어 MPEG-2(MPEG: Moving Picture Experts Group)나 H.264 등의 형식에 의한 파일 생성을 행하는 것이 가능해진다. 또한 정지 화상 파일로서 JPEG(Joint Photographic Experts Group), TIFF(Tagged Image File Format), GIF(Graphics Interchange Format) 등의 형식의 파일 생성을 행하는 것도 생각할 수 있다.

[0097] 센서 내 제어부(43)는, 예를 들어 CPU, ROM, RAM 등을 갖고 구성된 마이크로컴퓨터를 구비하여 구성되며, 이미지 센서(30)의 동작을 통괄적으로 제어한다. 예를 들어 센서 내 제어부(43)는, 촬상부(41)에 대한 지시를 행하여 촬상 동작의 실행 제어를 행한다. 또한, AI 처리부(44)나 화상 신호 처리부(42), 컴퓨터 비전 처리부(46)에 대해서도 처리의 실행 제어를 행한다.

[0098] 또한, 센서 내 제어부(43)는, 후술하는 바와 같이 서버 장치(1)로부터 이미지 센서(30)에 AI 모델(AI 모델의 설정 데이터)이 디플로이되었을 때, 해당 AI 모델이 AI 처리부(44)에 설정되도록 하는 처리를 행한다. 즉, AI 처리부(44)가 해당 AI 모델을 사용한 AI 처리를 실행 가능하게 되도록, AI 처리부(44)에 해당 AI 모델을 세팅하는 처리이다.

[0099] AI 처리부(44)는, 예를 들어 CPU나 FPGA(Field Programmable Gate Array), DSP(Digital Signal Processor) 등, 프로그래머블한 연산 처리 장치를 갖고 구성되며, 촬상 화상에 대하여 AI 처리를 행한다.

[0100] AI 처리부(44)에 의한 AI 처리로서는, 화상 인식 처리를 들 수 있다. 여기에서 말하는 화상 인식 처리란, 화상 내용의 인식을 행하는 처리를 넓게 의미하는 것이며, 예를 들어 피사체의 종별(예를 들어 사람, 동물, 차, 건물 등)의 인식 처리나, 피사체의 존재 유무나 범위의 인식 처리(소위 물체 검출 처리) 등을 들 수 있다.

[0101] AI 처리부(44)에 의한 AI 처리의 기능은, AI 처리부(44)에 설정하는 AI 모델(AI 처리의 알고리즘)을 변경함으로써 전환하는 것이 가능하다.

[0102] 메모리부(45)는, 불휘발성 메모리에 의해 구성되며, AI 처리부(44)에 의한 AI 처리를 행하는 데 있어서 필요한 데이터의 기억에 사용된다. 구체적으로는, AI 모델의 설정 데이터(예를 들어, 뉴럴 네트워크에 컨볼루션 연산에서 사용하는 각종 가중 계수나, 뉴럴 네트워크의 구조를 나타내는 데이터 등)는, 메모리부(45)에 기억된다.

[0103] 또한, 본 예에 있어서, 메모리부(45)는, 화상 신호 처리부(42)에서 처리된 촬상 화상 데이터의 보유에도 사용된다.

[0104] 컴퓨터 비전 처리부(46)는, 촬상 화상 데이터에 대한 화상 처리로서, 룰 베이스에 의한 화상 처리를 실시한다. 여기에서의 룰 베이스에 의한 화상 처리로서는, 예를 들어 초해상 처리 등을 들 수 있다.

[0105] 통신 인터페이스(47)는, 이미지 센서(30) 외부에 있어서의 제어부(33)나 메모리부(34) 등, 버스(37)를 통해 접속된 각 부와의 사이에서 통신을 행하는 인터페이스이다. 예를 들어 통신 인터페이스(47)는, 센서 내 제어부(43)의 제어에 기초하여, AI 처리부(44)가 사용하는 AI 모델 등을 외부로부터 취득하기 위한 통신을 행한다.

[0106] 또한, 통신 인터페이스(47)를 통해, AI 처리부(44)에 의한 AI 처리의 처리 결과 정보가 이미지 센서(30)의 외부에 출력된다.

[0107] 여기서, 이미지 센서(30)에 있어서, 센서 내 제어부(43)는, 통신 인터페이스(47), 통신부(35), 및 포그 서버(4)를 통해, 서버 장치(1)와의 사이에서 데이터 통신을 행하는 것이 가능하게 되어 있다. 이에 의해 센서 내 제어부(43)는, 후술하는 바와 같이 AI 모델의 설정 데이터 등의 각종 데이터를 서버 장치(1)로부터 수신하거나, AI 처리부(44)에 의한 처리 결과 정보 등의 각종 데이터를 서버 장치(1)에 대하여 송신하거나 하는 것이 가능하게 된다.

[0108] 그런데, 상기 설명과 같이 본 예에서는, 이미지 센서(30)가, 화소 어레이부를 갖는 촬상부(41) 이외에, AI 처리를 행하는 AI 처리부(44)나, 컴퓨터 장치로서의 센서 내 제어부(43) 등을 갖는 것으로 되어 있다. 이와 같은 이미지 센서(30)의 구조예를 도 4를 참조하여 설명해 둔다. 또한, 도 4에 도시한 구조는 어디까지나 일례이며, 물론 다른 구조를 채용하는 것도 가능하다.

[0109] 도 4에 도시한 바와 같이, 본 예에 있어서의 이미지 센서(30)는, 다이 D1과 다이 D2의 2개의 다이를 적층한 2층 구조(적층 구조)를 갖는 것으로 된다. 구체적으로, 본 예에 있어서의 이미지 센서(30)는, 다이 D1과 다이 D2를 접합한 1칩의 반도체 장치로서 구성되어 있다. 다이 D1은, 촬상부(41)가 형성된 다이이며, 다이 D2는, 화상 신호 처리부(42), 센서 내 제어부(43), AI 처리부(44), 메모리부(45), 컴퓨터 비전 처리부(46), 및 통신 인터페이스(47)를 구비한 다이이다.

[0110] 다이 D1과 다이 D2는, 예를 들어 Cu-Cu 접합 등의 칩간 접합 기술에 의해 물리적 또한 전기적으로 접속되어 있다.

[0111] <3. 실시 형태로서의 적합화 방법>

[0112] 우선은 도 5를 참조하여, 본 실시 형태의 카메라(3)에서 사용되는 AI 모델(이하 「AI 모델(50)」로 표기)의 개요를 설명해 둔다.

[0113] 본 실시 형태에 있어서, AI 모델(50)로서는, 뉴럴 네트워크, 구체적으로는 DNN(Deep Neural Network)을 갖는 AI 모델을 상정하고 있다. 보다 구체적으로는, 이미지 센서(30)에 있어서의 촬상부(41)에서 얻어지는 촬상 화상(본 예에서는 RGB 화상)을 입력 데이터로 하고, 화상 인식 처리로서의 추론 처리를 행하는 AI 모델로 된다.

[0114] 도면 중, 후 처리부(55)는, AI 모델(50)의 출력 데이터에 기초하여 화상 인식 처리의 결과를 나타내는 인식 결과 정보를 디코드한다. 본 예에서는, 화상 인식 처리로서는 인물로서의 타깃 피사체의 연령이나 성별 등과 같은 속성을 인식하는 처리로 되어 있고, 이것에 따라 후 처리부(55)는, AI 모델로부터의 출력 데이터에 기초하여, 이들 연령이나 성별과 같은 속성의 인식 결과를 나타내는 정보를 디코드한다.

[0115] 또한, 본 예에 있어서 AI 모델(50)에서는, 인식 결과에 대한 스코어(우도)를 나타내는 값이 산출되고, 후 처리부(55)는 이 값을 디코드하여, 인식 결과 정보로서 해당 스코어를 포함하는 정보를 얻는다.

[0116] 본 실시 형태에서는, 상기와 같은 화상 인식 처리로서의 추론 처리를 행하는 AI 모델(50)에 대하여, 카메라(3)에 있어서의 이미지 센서(30)에 대해, 우선은 초기 AI 모델(51)로서의 AI 모델(50)을 디플로이하는 전제로 된다.

[0117] 도 6은 초기 AI 모델(51)의 설명도이다.

[0118] 초기 AI 모델(51)로서는, 이미지 센서(30)의 다양한 사용 환경 조건에 대응 가능하게 되도록 기계 학습된, 범용적인 AI 모델(50)을 준비해 둔다.

[0119] 도 6의 A는 초기 AI 모델(51)을 생성하기 위한 기계 학습에 대한 설명도이다. 이 경우의 기계 학습에

있어서는, 화상 인식 처리로서의 추론 처리를 실현하기 위한 소정의 네트워크 구조를 갖는 DNN 네트워크(50n)를 준비함과 함께, 학습용 입력 데이터로서의 복수의 화상 데이터와, 그것들 화상 데이터마다 화상 인식 결과의 정해 정보를 나타내는 라벨 데이터로 이루어지는 학습용 데이터 세트를 준비한다.

[0120] 이 경우, 학습용 입력 데이터로서 준비하는 화상 데이터에 대해서는, 학습 후의 초기 AI 모델(51)이 범용성을 갖도록, 이미지 센서(30)의 사용 환경 조건으로서 상정되는 각 조건에 대응한 화상 데이터를 준비한다. 즉, 각각의 사용 환경 조건에서 촬상이 행해진 경우에 얻어지는 촬상 화상을 준비한다. 예를 들어, 상정하는 사용 환경 조건이 예를 들어 일본, 미국, 유럽 등의 사용 지역에 관한 조건인 경우에는, 그것들 사용 지역마다, 그 지역에서 촬상이 행해진 경우에 얻어지는 촬상 화상을 준비한다.

[0121] 상기와 같은 상정되는 각 사용 환경 조건에 대응한 화상 데이터와, 그 라벨 데이터를 포함하는 학습용 데이터 세트를 사용하여, DNN 네트워크(50n)에 대하여 기계 학습을 행함으로써, 범용성을 갖는 초기 AI 모델(51)을 생성한다(도 6의 B).

[0122] 여기서, 이와 같이 생성한 초기 AI 모델(51)을, 이미지 센서(30)의 사용 환경 조건에 적합한 모델로 커스터마이즈(적합화)하는 것을 생각한다.

[0123] 이 커스터마이즈에 대해서는, 초기 AI 모델(51)의 네트워크 전체를 재학습한다고 하는 방법으로 실현하는 것을 생각할 수 있다. 그러나 전술한 바와 같이, 네트워크 전체를 재학습하기 위해서는 필요한 학습 횟수가 증대화되어, 커스터마이즈에 요하는 시간의 증대화를 초래하고, 또한, 재학습을 위해 유저가 서버 장치(1)에 업로드해야 할 학습용 화상 데이터의 수도 증대되어, 커스터마이즈에 요하는 통신 데이터양의 증대화를 초래하게 된다.

[0124] 따라서, 본 실시 형태에서는, 초기 AI 모델을 소정 위치에서 분단하고, 소정 중간층보다도 후단의 네트워크인 후단 네트워크만을 커스터마이즈의 대상으로 한다고 하는 방법을 채용한다.

[0125] 도 7은 초기 AI 모델(51)의 분단, 및 전단 네트워크와 후단 네트워크의 설명도이다.

[0126] 초기 AI 모델(51)이 갖는 복수의 중간층에 대하여, 소정의 층간 위치를 분단 위치 Dv로 하여 초기 AI 모델(51)을 분단한다(도 7의 A). 이 분단 위치 Dv보다도 전단측의 네트워크를 「전단 네트워크」, 후단측의 네트워크를 「후단 네트워크」로 한다(도 7의 B).

[0127] 이 경우, 전단 네트워크는, 분단 위치 Dv의 직전의 중간층에서 얻어지는 특징 맵을 출력한다. 이와 같이 전단 네트워크가 출력하게 되는, 분단 위치 Dv의 직전의 중간층에서 얻어지는 특징 맵을 이하 「중간 특징 맵 IRM」으로 표기한다.

[0128] 이후에서 설명하는 바와 같이, 본 실시 형태에서는, 후단 네트워크의 커스터마이즈 시에, 이미지 센서(30) 측(에지측)으로부터 서버 장치(1) 측(클라우드측)에 대하여 중간 특징 맵 IRM을 송신하는 것으로 된다.

[0129] 중간 특징 맵 IRM은, 그대로는 개인 정보를 특정하는 것이 곤란한 데이터로 된다. 이 때문에, 상기와 같이 이미지 센서(30) 측으로부터 서버 장치(1) 측에 중간 특징 맵 IRM을 송신하는 것을 요하는 경우라도, 개인 정보의 유출 가능성의 저감을 도모할 수 있다. 즉, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서, 개인 정보의 유출 가능성의 저감이 도모된다.

[0130] 본 실시 형태에 있어서, 분단 위치 Dv로서는, 초기 AI 모델(51)에 있어서의 중간층의 층간 위치 중, 2번째 이후의 층간 위치로 정하는 것으로 하고 있다. 환언하면, 중간 특징 맵 IRM은, 초기 AI 모델(51)에 있어서의 2층째 이후의 중간층의 출력 데이터로 되는 것이다.

[0131] 2층째 이후의 중간층에 있어서는, 출력 데이터로서, 1층째의 중간층의 출력 데이터보다도 개인 정보의 특정이 곤란해지는 데이터가 얻어진다고 할 수 있다(입력 데이터가 보다 차원 압축되기 때문에).

[0132] 따라서, 상기와 같은 분단 위치 Dv의 설정에 의해, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0133] 본 실시 형태에서는, 초기 AI 모델(51)의 커스터마이즈 방법으로서, 후단 네트워크의 후보를 복수 준비해 두고, 그것들 후보로서의 후단 네트워크 중에서 하나의 후단 네트워크를, 대상으로 하는 이미지 센서(30)로부터 취득한 중간 특징 맵 IRM에 기초하여 선택한다고 하는 방법을 채용한다.

[0134] 도 8은 후단 네트워크의 후보의 생성 방법예에 대한 설명도이다.

[0135] 후단 네트워크의 복수의 후보는, 초기 AI 모델(51)에 있어서의 후단 네트워크만을, 각각 다른 학습용 데이터 세트를 사용하여 재학습함으로써 생성한다.

[0136] 도시와 같이 본 예에서는, 재학습을 위한 학습용 데이터 세트로서, N종류의 학습용 데이터 세트(제1 학습용 데이터 세트로부터 제N 학습용 데이터 세트)를 준비하는 것으로 하고, 초기 AI 모델(51)에 있어서의 후단 네트워크의 재학습으로서, 이들 N종류의 학습용 데이터 세트 중 각각 대응하는 하나를 사용한 N가지의 재학습을 행하여, N종류의 후단 네트워크를 생성한다. 이때, 후단 네트워크만의 재학습은, 전단 네트워크에 있어서의 가중계수를 고정으로 하여 행할 수 있다.

[0137] 이 경우, 각 학습용 데이터 세트로서는, 그것들 학습용 데이터 세트간에서 학습용 입력 데이터로서 포함되는 화상 데이터의 종류가 다른 것이 사용된다. 구체적으로는, 각각이 다른 환경에서 촬상된 화상 데이터를 학습용 입력 데이터로서 포함하는 학습용 데이터 세트가 사용되는 것이다. 예를 들어, 제1 학습용 데이터 세트에는, 학습용 입력 데이터로서 제1 환경에서 촬상된 복수의 화상 데이터가 포함되고, 제2 학습용 데이터 세트에는, 학습용 입력 데이터로서 제1 환경과는 다른 제2 환경에서 촬상된 복수의 화상 데이터가 포함되는 식이다.

[0138] 이와 같은 N종류의 학습용 데이터 세트를 사용한 후단 네트워크의 재학습이 행해짐으로써, 도 9에 도시한 바와 같이, 각각이 N종류의 환경 중 하나에 적합한, N개의 후단 네트워크의 후보가 생성된다.

[0139] 본 실시 형태에 있어서, 이와 같이 생성된 N종류의 후단 네트워크의 후보는, 서버 장치(1)에 있어서의 기억부(19) 등, 서버 장치(1)의 CPU(11)가 읽어내기 가능한 기억 장치에 기억된다.

[0140] 서버 장치(1)는, 이들 후보로서의 후단 네트워크 중에서, 대상으로 하는 이미지 센서(30)로부터 취득한 중간 특징 맵 IRM에 기초하여, 해당 이미지 센서(30)의 사용 환경 조건에 적합한 하나의 후단 네트워크를 선택한다.

[0141] 도 10은 이와 같은 후단 네트워크의 선택에 관한 서버 장치(1), 이미지 센서(30) 각각의 기능을 설명하기 위한 도면이다.

[0142] 도시와 같이 서버 장치(1)는 후단 네트워크 취득부 F11로서의 기능을 갖고, 이미지 센서(30)는 송신 처리부 F31로서의 기능을 갖는다. 본 예에 있어서, 후단 네트워크 취득부 F11로서의 기능은, 서버 장치(1)의 CPU(11)에 의한 소프트웨어 처리로 실현되는 기능으로 되고, 송신 처리부 F31로서의 기능은 이미지 센서(30)에 있어서의 센서 내 제어부(43)의 소프트웨어 처리로 실현되는 기능으로 된다.

[0143] 이미지 센서(30)에 있어서, 송신 처리부 F31은, 초기 AI 모델(51)에 대해 입력 데이터를 부여하였을 때 얻어지는 중간 특징 맵 IRM을 외부에 송신하는 처리를 행한다. 구체적으로는, 촬상부(41)에 의해 얻어지는 촬상 화상(즉 이미지 센서(30)의 사용 환경에 있어서 촬상된 화상)을 초기 AI 모델(51)의 입력 데이터로서 부여하였을 때 얻어지는 중간 특징 맵 IRM을 서버 장치(1)에 송신하기 위한 처리를 행하는 것이다. 이 경우, 송신 처리부 F31(센서 내 제어부(43))은, 중간 특징 맵 IRM을 통신 인터페이스(47)를 통해 이미지 센서(30) 외로 출력하는 처리를 행함과 함께, 예를 들어 제어부(33)에 대해, 해당 중간 특징 맵 IRM을 서버 장치(1)에 대하여 송신하는 처리를 실행하도록 지시를 행한다. 이에 의해 해당 중간 특징 맵 IRM은, 카메라(3)에 있어서의 통신부(35)를 경유하여, 포그 서버(4)를 통해 서버 장치(1)에 송신된다.

[0144] 서버 장치(1)에 있어서, 후단 네트워크 취득부 F11은, 이와 같이 이미지 센서(30) 측으로부터 송신된 중간 특징 맵 IRM을 입력하고, 입력한 중간 특징 맵 IRM에 기초하여, 복수의 후단 네트워크의 후보 중에서 하나의 후단 네트워크를 선택한다.

[0145] 이때, 복수의 후단 네트워크의 후보로부터, 중간 특징 맵 IRM에 기초하여 하나의 후단 네트워크를 선택하는 방법에 대해서는 다양한 방법을 생각할 수 있다.

[0146] 일례로서는, 예를 들어 도 11에 도시한 바와 같이, 후단 네트워크의 후보마다 중간 특징 맵 IRM을 입력 데이터로서 부여하여 추론 처리를 실행시키고, 후 처리부(55)에 의해 계산되는 추론 결과에 대한 스코어에 기초하여, 하나의 후단 네트워크를 선택한다고 하는 방법을 들 수 있다. 구체적으로는, 가장 스코어가 양호한 후단 네트워크를, 대상으로 하는 이미지 센서(30)의 사용 환경에 적합한 후단 네트워크로서 선택하는 것이다.

[0147] 또한, 후단 네트워크의 후보수가 많은 경우 등에는, 예를 들어 이분 탐색법 등을 채용하여 선택에 관한 처리량의 저감을 도모하는 것을 생각할 수 있다.

[0148] 혹은, N개의 후보에 대하여 소정의 순번으로 중간 특징 맵 IRM을 입력 데이터로서 부여하여 스코어의 계산을 행하고, 스코어가 소정의 역치 이상이 된 후단 네트워크가 검출된 시점에서, 그 후단 네트워크를 적합한 후단 네트워크로서 선택한다고 하는 것도 생각할 수 있다.

[0149] 또한, 스코어에 기초하는 선택으로 하는 것이 아니라, 중간 특징 맵 IRM의 해석 결과에 기초하여, 적합한 후단

네트워크를 선택한다고 하는 것도 생각할 수 있다. 예를 들어, 미리 중간 특징 맵 IRM에 있어서의 값의 분포와 그것에 적합한 후단 네트워크의 대응 관계를 구해 두고, 해당 분포의 해석 결과에 기초하여 적합한 후단 네트워크를 선택한다고 하는 것을 생각할 수 있다.

[0150] 혹은, 중간 특징 맵 IRM에 기초하는 후단 네트워크의 선택은, AI를 사용하여 행하는 것도 생각할 수 있다. 구체적으로는, 예를 들어 중간 특징 맵 IRM을 학습용 입력 데이터로 하고, 그 중간 특징 맵 IRM에 대하여 선택되어야 할 후단 네트워크의 정해 정보를 교사 데이터로 하여 기계 학습한 AI 모델을 사용하는 것 등을 생각할 수 있다.

[0151] 도 12는 선택한 후단 네트워크의 디플로이에 관한 서버 장치(1), 이미지 센서(30) 각각이 갖는 기능을 설명하기 위한 도면이다.

[0152] 선택한 후단 네트워크의 디플로이에 관한 기능으로서, 도시와 같이 서버 장치(1)는 송신 처리부 F12로서의 기능을 갖고, 이미지 센서(30)는 수신 처리부 F32로서의 기능을 갖고 있다. 또한 본 예에 있어서, 이들 송신 처리부 F12, 수신 처리부 F32의 기능으로서도, 각각, 서버 장치(1)의 CPU(11)에 의한 소프트웨어 처리, 이미지 센서(30)에 있어서의 센서 내 제어부(43)의 소프트웨어 처리로 실현되는 기능으로 된다.

[0153] 서버 장치(1)에 있어서, 송신 처리부 F12는, 후단 네트워크 취득부 F11이 선택한 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행한다. 구체적으로는, 후단 네트워크 취득부 F11이 선택한 후단 네트워크의 설정 데이터를 통신부(20)를 통해 이미지 센서(30)에 송신하기 위한 처리를 행한다. 이 처리가 행해짐으로써, 당해 후단 네트워크의 설정 데이터는 포그 서버(4)를 경유하여 이미지 센서(30)에 송신되는 것이 된다.

[0154] 이미지 센서(30)에 있어서, 수신 처리부 F32는, 송신 처리부 F12에 의해 송신된 후단 네트워크의 설정 데이터를 수신하는 처리를 행한다. 이것은, 앞서 설명한 송신 처리부 F31(도 10 참조)이 송신한 중간 특징 맵 IRM에 기초하여, 서버 장치(1)가 후단 네트워크의 복수의 후보 중에서 선택한 하나의 후단 네트워크의 설정 데이터를 수신하는 처리를 행하는 것이라고 환언할 수 있다.

[0155] 도시에 의한 설명은 생략하지만, 이미지 센서(30)에 있어서는, 이와 같이 수신 처리부 F32에 의해 수신된 설정 데이터에 기초하여, AI 처리부(44)에 설정된 초기 AI 모델(51)에 있어서의 후단 네트워크의 갱신이 행해진다. 이에 의해 이후, AI 처리부(44)에 있어서는, 갱신 후의 후단 네트워크, 즉, 이미지 센서(30)의 사용 환경 조건에 적합한 것으로서 서버 장치(1) 측에서 선택된 후단 네트워크를 사용한 추론 처리가 행해지는 것이 된다.

[0156] 여기서, 상기에서는 전단 네트워크와 후단 네트워크의 분단 위치 Dv에 대하여, 2번째 이후의 층간 위치로 설정하는 것을 언급하였다.

[0157] 도 13은 초기 AI 모델(51)에 대하여, 2개의 다른 분단 위치 Dv에서 얻어지는 중간 특징 맵 IRM의 가시화 이미지를 예시하고 있다. 구체적으로는, 분단 위치 Dv를 1번째와 2번째의 중간층의 층간 위치로 하였을 때의 중간 특징 맵 IRM의 가시화 이미지와, 분단 위치 Dv를 3번째와 4번째의 중간층의 층간 위치로 하였을 때의 중간 특징 맵 IRM의 가시화 이미지를 예시하고 있다.

[0158] 이 도 13으로부터, 1번째의 중간층보다도 후단측의 중간층에서는, 중간 특징 맵 IRM의 화상 사이즈가 작아지는 경향이 되고, 또한 가시화하였을 때의 화상 내용으로서도 개인 정보를 특정하기 어려워지는 것을 알 수 있다.

[0159] 이와 같은 중간 특징 맵 IRM의 특성을 감안하여, 분단 위치 Dv에 대해서는, 이하와 같이 설정하는 것도 생각할 수 있다.

[0160] 하나는, 중간 특징 맵 IRM을 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 소정 화소수 미만이 되도록 정한다고 하는 방법을 생각할 수 있다.

[0161] 도 14는 초기 AI 모델(51)에 대한 입력 데이터로서의 촬상 화상과, 해당 촬상 화상을 입력 데이터로서 부여하였을 때 초기 AI 모델(51)에 있어서 얻어지는 중간 특징 맵 IRM의 가시화 이미지를 대비하여 나타내고 있다. 여기에서는, 촬상 화상에는 개인 정보에 관한 화상 영역으로서 사람의 얼굴이 찍힌 화상 영역(「화상 영역 Ar1」로 함)이 존재하고 있는 예를 나타내고 있다. 이 경우, 중간 특징 맵 IRM의 가시화 이미지에는, 도면 중에 나타내는 화상 영역 Ar2와 같이, 이 화상 영역 Ar1에 대응한, 개인 정보에 관한 화상 영역이 존재하게 되지만, 이 경우에 있어서, 해당 화상 영역 Ar2의 화소수가, 소정 화소수 미만이 되도록 분단 위치 Dv를 정하는 것이다.

[0162] 이때, 「소정 화소수」는, 중간 특징 맵 IRM을 가시화하였을 때 개인 정보의 특정이 곤란해지도록 하는 화소수를 정하면 되고, 실험적으로는, 예를 들어 12×12화소에 대응한 144화소로 하는 것이 바람직하고, 또한, 8×8화

소에 대응한 64화소로 하는 것이 보다 바람직한 것이 된다.

[0163] 또한, 분단 위치 Dv의 다른 설정 방법으로서는, 중간 특징 맵 IRM이, 대상으로 하는 AI 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터가 되도록 정하는 것을 생각할 수 있다.

[0164] 여기서, 자기 부호화 학습이란, 오토 인코더 작성을 위한 사전 학습이며, 구체적으로는, 출력 데이터를 입력 데이터와 일치시키는 비지도 학습을 의미한다.

[0165] 도 15는 해당 다른 설정 방법에 대한 설명도이다.

[0166] 이 경우에는 도 15의 A에 도시한 바와 같이, 우선, 초기 AI 모델(51)에 사용하는 DNN 네트워크(50n)에 대하여, 자기 부호화 학습을 행하여, 오토 인코더(60)를 생성한다. 그리고, 도 15의 B에 도시한 바와 같이, 이 오토 인코더(60)의 디코드부(60a)에 중간 특징 맵 IRM을 입력하여, 해당 중간 특징 맵 IRM을 디코드할 수 있었는지 여부를 판정한다. 이 판정은, 해당 중간 특징 맵 IRM을 얻음에 있어서 입력 데이터로서 사용한 화상 데이터와, 디코드부(60a)의 출력 데이터를 비교한 결과에 기초하여 행하는 것을 생각할 수 있다.

[0167] 예를 들어, 초기 AI 모델(51)에 대하여, 분단 위치 Dv를 차례로 후단측으로 천이시켜 가고, 분단 위치 Dv마다, 중간 특징 맵 IRM에 대하여 상기와 같은 디코드부(60a)를 사용한 판정을 행한다. 이때, 중간 특징 맵 IRM의 디코드가 불가능하게 되었다고 판정되었을 때의 분단 위치 Dv를, 채용하는 분단 위치 Dv로서 정하는 것을 생각할 수 있다.

[0168] <4. 처리 수순>

[0169] 도 16은 상기에 의해 설명한 실시 형태로서의 적합화 방법을 실현하기 위한 처리 수순예를 나타낸 흐름도이다.

[0170] 도 16에 있어서, 「서버 장치」로서 나타내는 처리는, 서버 장치(1)에 있어서의 CPU(11)가 예를 들어 ROM(12) 등의 소정의 기억 장치에 기억된 프로그램에 기초하여 실행하는 것이며, 「이미지 센서」로서 나타내는 처리는 이미지 센서(30)에 있어서의 센서 내 제어부(43)의 CPU가 센서 내 제어부(43)의 ROM 등의 소정의 기억 장치에 기억된 프로그램에 기초하여 실행한다.

[0171] 우선, 서버 장치(1) 측에 있어서, CPU(11)는 스텝 S101에서, 대상 에지의 지시, 즉, 후단 네트워크의 적합화 대상으로 하는 이미지 센서(30)의 지시가 행해질 때까지 대기한다. 이 대상 에지의 지시는, 본 예에서는, 유저가 유저 단말기(2)에 대하여 행한 조작 입력에 기초하여, 유저 단말기(2)가 서버 장치(1)에 대하여 행하는 것으로 된다.

[0172] 대상 에지의 지시가 있는 경우, CPU(11)는 스텝 S102에서, 대상이 되는 이미지 센서(30)에 중간 특징 맵 생성 동작의 실행 지시를 행한다. 즉, 중간 특징 맵 IRM의 생성 동작의 실행 지시를 행하는 것이다.

[0173] 이미지 센서(30) 측에 있어서, 센서 내 제어부(43)는, 이와 같은 중간 특징 맵 생성 동작의 실행 지시를, 스텝 S201에 있어서 대기하고 있고, 해당 실행 지시가 있는 경우에는, 스텝 S202에서 생성 동작 실행 처리를 행한다. 즉, 촬상부(41)에 의한 촬상 동작을 실행시킴과 함께, 해당 촬상 동작에 의해 얻어진 촬상 화상을 AI 처리부(44)에 있어서의 초기 AI 모델(51)의 입력 데이터로서 부여함으로써, 중간 특징 맵 IRM을 생성시키는 것이다.

[0174] 스텝 S202에 이은 스텝 S203에서 센서 내 제어부(43)는, 중간 특징 맵 IRM을 서버 장치(1)에 송신하는 처리를 행한다. 이것은, 전술한 송신 처리부 F31의 처리에 상당하는 것이다.

[0175] 서버 장치(1) 측에 있어서, CPU(11)는, 이미지 센서(30) 측으로부터의 중간 특징 맵 IRM의 수신을 스텝 S103에서 대기하고 있고, 중간 특징 맵 IRM을 수신한 경우에는, 스텝 S104에서, 수신한 중간 특징 맵 IRM에 기초하는 후단 네트워크의 선택 처리를 실행한다. 구체적으로는, 앞의 후단 네트워크 취득부 F11로서 설명한 바와 같이, 복수의 후단 네트워크의 후보 중에서, 수신한 중간 특징 맵 IRM에 기초하여 하나의 후단 네트워크를 선택하는 처리를 행하는 것이다. 또한, 수신한 중간 특징 맵 IRM에 기초하는 후단 네트워크의 선택 방법의 구체예에 대해서는 이미 설명하였기 때문에 중복 설명을 피한다.

[0176] 스텝 S104에 이은 스텝 S105에서 CPU(11)는, 선택한 후단 네트워크의 설정 데이터를 이미지 센서(30)에 송신하는 처리(즉 전술한 송신 처리부 F12에 상당하는 처리)를 행하고, 도 16에 나타내는 일련의 처리를 종료한다.

[0177] 이미지 센서(30) 측에 있어서, 센서 내 제어부(43)는, 스텝 S105에서 송신된 설정 데이터의 수신을 스텝 S204에서 대기하고 있고, 해당 설정 데이터를 수신한 경우에는, 스텝 S205에서, 해당 설정 데이터에 기초하는 후단 네

트워크의 설정 처리를 행한다. 즉, AI 처리부(44)에 있어서의 초기 AI 모델(51)의 후단 네트워크를, 해당 설정 데이터에 기초하여 갱신하는 것이다.

[0178] 센서 내 제어부(43)는, 스텝 S205의 처리를 실행한 것에 따라서 도 16에 도시한 일련의 처리를 종료한다.

[0179] 또한, 상기에서는, 서버 장치(1)가 선택한 후단 네트워크를 그대로 에지측에 송신하는 예를 들었지만, 선택한 후단 네트워크에 대하여, 파인 튜닝으로서의 재학습을 행하고, 해당 재학습 후의 후단 네트워크를 에지측에 송신한다고 하는 것도 생각할 수 있다.

[0180] <5. 후단 네트워크의 소형화에 대하여>

[0181] 여기서, 본 실시 형태와 같이 이미지 센서(30)의 AI 처리부(44)에 있어서 AI 처리를 행하는 구성에 있어서는, AI 처리에 사용할 수 있는 하드웨어 리소스가 부족한 경우가 상정되고, 이것에 수반하여, AI 처리에 사용하는 뉴럴 네트워크에 대해서도 사이즈적인 제한을 받는 것이 상정된다.

[0182] 이 때문에, 후단 네트워크에 대하여 소형화가 도모되는 것이 바람직하다.

[0183] 단, 네트워크가 소형이 되면, 화상 인식 정밀도의 저하 등, 추론 처리의 정밀도 저하를 초래할 우려가 있다.

[0184] 따라서, 후단 네트워크에 대하여, 네트워크 사이즈의 소형화를 추론 처리의 정밀도 저하의 억제를 도모하면서 실현하기 위한 방법으로서, 이하와 같은 액티브 러닝을 적용한 방법을 채용하는 것을 생각할 수 있다.

[0185] 도 17은 액티브 러닝을 적용한 후단 네트워크의 생성 방법에 대한 설명도이다.

[0186] 이 경우에는, 후단 네트워크로서 소형 네트워크를 사용하는 것을 전제로 한다.

[0187] 초기 AI 모델(51)에 있어서의 전단 네트워크와, 소형의 후단 네트워크를 조합한 학습기의 기계 학습으로서, 이미지 센서(30)의 사용 환경으로서 상정하는 N종의 환경마다, 각각 대응하는 환경 화상을 학습용 입력 데이터로서 사용한 기계 학습을 행한다. 이에 의해, 소형의 후단 네트워크로서, N개의 환경 대응 학습 완료 네트워크가 얻어진다.

[0188] 단, 이 경우의 후단 네트워크는 소형 네트워크이기 때문에, 상기와 같은 환경 대응 학습이 행해졌다고 해도, 높은 추론 성능을 기대할 수는 없다.

[0189] 따라서, 소형이라도 환경에 특화된 후단 네트워크를 생성할 수 있도록, 액티브 러닝에 의한 기계 학습을 행한다. 구체적으로는, 상기한 환경 대응 학습을 행하였을 때 후 처리부(55)에서 산출된 스코어를 참조하여, 스코어가 중 정도(예를 들어, 약 0.5 내지 약 0.7의 범위 등)였던 입력 화상 데이터를, 어노테이터에 제시한다(도 17의 A 참조). 도면 중에서는, 이와 같이 스코어가 중 정도가 되는 종별의 화상 데이터를 「종별 B」의 화상 데이터로서 나타내고 있다.

[0190] 이 종별 B에 해당하는 화상 데이터를 학습용 입력 데이터로서 포함하는 재학습용 데이터 세트를 사용하여, 소형 네트워크에 의한 후단 네트워크를 포함하는 네트워크 전체에 대하여, 재학습을 행한다(도 17의 B 참조).

[0191] 상기와 같은 액티브 러닝에 의한 후단 네트워크 생성을, 상정하는 N종의 환경마다 각각 행한다.

[0192] 이에 의해, 상정하는 N종의 환경마다의 후단 네트워크에 대하여, 소형이라도 환경에 특화된 것을 효율적으로 생성할 수 있다.

[0193] 이 경우의 서버 장치(1)에 있어서의 후단 네트워크 취득부 F11은, 상기와 같은 액티브 러닝에 의해 생성된 복수의 후단 네트워크의 후보 중에서, 대상으로 하는 이미지 센서(30)로부터 입력된 중간 특징 맵 IRM에 기초하여, 하나의 후단 네트워크를 선택한다.

[0194] 또한, 후단 네트워크의 소형의 방법으로서는, 증류(지식 증류: Knowledge Distillation)를 들 수 있다.

[0195] 구체적인 방법의 하나로서는, 범용 또한 대형의 마스터 AI를 교사 모델로서 준비해 두고, 그것에 대하여 증류를 행한다고 하는 방법을 들 수 있다.

[0196] 구체적으로는, 도 18에 도시한 바와 같이, 초기 AI 모델(51)에 있어서의 후단 네트워크를, 범용 또한 대형의 마스터 AI로서의, 교사 모델로서 사용하는 것으로 하고, 해당 교사 모델에 대한 증류를 행하여, 소형 후단 네트워크를 생성한다.

[0197] 이 경우의 증류에 있어서는, 네트워크의 소형화뿐만 아니라, 대상으로 하는 이미지 센서(30)의 사용 환경 조건

에 따른 모델 적합화도 행해지도록 한다. 공지와 같이 증류는, 기본적으로는, 동일한 입력 데이터에 대한 교사 모델의 추론 결과를 소프트 타깃으로서 사용하여, 학생 모델의 추론 결과가 해당 소프트 타깃에 가까워지도록 학생 모델의 학습을 행하는 것이 되지만, 이때, 교사 모델, 학생 모델에 공통의 입력 데이터로서, 대상으로 하는 이미지 센서(30) 측으로부터 입력한 중간 특징 맵 IRM을 사용함으로써 네트워크의 소형화와 함께, 해당 이미지 센서(30)의 사용 환경 조건에 따른 모델 적합화도 행해지도록 한다.

[0198] 이 도 18에 도시한 바와 같은 증류를 행하는 경우에는, 후단 네트워크의 복수의 후보를 미리 생성해 둘 필요는 없다. 즉, 이미지 센서(30) 측의 사용 환경 조건에 적합한 후단 네트워크를 디플로이함에 있어서, 미리 준비된 복수의 후단 네트워크의 후보로부터 하나의 후단 네트워크를 선택한다고 하는 것은 행하지 않는다.

[0199] 상기 설명으로부터 이해되는 바와 같이, 이 경우에 있어서의 후단 네트워크 취득부 F11은, 베이스가 되는 학습 완료 네트워크와 중간 특징 맵 IRM에 기초하여, 하나의 후단 네트워크를 생성하는 처리를 행하게 된다. 구체적으로, 이 경우에 있어서의 후단 네트워크 취득부 F11은, 초기 AI 모델(51)의 후단 네트워크를 교사 모델로 하여, 이미지 센서(30) 측으로부터 입력된 중간 특징 맵 IRM에 기초하는 증류를 행함으로써, 이미지 센서(30) 측의 사용 환경 조건에 적합한 소형 후단 네트워크를 생성하는 것이다.

[0200] 또한, 증류를 적용한 소형 후단 네트워크 생성의 다른 방법으로서는, 범용의 후단 네트워크를 교사 모델로 하는 것이 아니라, 어느 정도 환경에 특화되도록 학습된 후단 네트워크를 교사 모델로서 사용하는 방법도 들 수 있다.

[0201] 도 19는 해당 다른 방법에 대한 설명도이다.

[0202] 이 경우에는, 증류에 있어서의 교사 모델의 후보로서, 각각이 다른 사용 환경에 적합하도록 학습된, 복수의 대형 후단 네트워크를 준비해 둔다(도 19의 A 참조).

[0203] 그리고, 이 경우에 있어서의 후단 네트워크 취득부 F11은, 이와 같이 준비된 복수의 대형 후단 네트워크의 후보 중에서, 대상으로 하는 이미지 센서(30) 측으로부터 입력한 중간 특징 맵 IRM에 기초하여, 대응하는 하나의 대형 후단 네트워크를 교사 모델로서 선택한다(도 19의 B 참조). 즉, 입력한 중간 특징 맵 IRM의 예를 들어 수치 분포 등으로부터, 이미지 센서(30)의 사용 환경을 추정하고, 해당 추정한 사용 환경에 대응하는 대형 후단 네트워크를 선택한다.

[0204] 그리고 나서, 이 경우의 후단 네트워크 취득부 F11은, 선택한 대형 후단 네트워크를 교사 모델로 한 증류로서, 교사 모델, 학생 모델의 입력 데이터를 이미지 센서(30) 측으로부터 입력한 중간 특징 맵 IRM으로 한 증류의 처리를 행함으로써, 해당 이미지 센서(30)의 사용 환경에 적합한 하나의 후단 네트워크를 생성한다(도 19의 C 참조).

[0205] 상기와 같은 다른 방법으로서의 증류를 행함으로써, 도 18에 도시한 바와 같이 범용의 대형 후단 네트워크를 교사 모델로 한 증류를 행하는 경우보다도, 폭넓은 사용 환경에 대응 가능하게 되도록 할 수 있다.

[0206] 또한, 상기에서는 후단 네트워크의 소형화의 예로서 액티브 러닝이나 증류를 행하는 예를 들었지만, 후단 네트워크의 소형화로서는, 가지치기(Pruning)나 양자화를 행하는 것도 생각할 수 있다.

[0207] 확인을 위해, 도 20의 흐름도에, 상기와 같은 증류에 의한 소형 후단 네트워크의 생성을 행하는 경우에 대응한 서버 장치(1), 이미지 센서(30)의 처리 수순예를 나타낸다.

[0208] 또한, 이미지 센서(30) 측의 처리 수순예에 대해서는 앞의 도 16에서 설명한 것과 마찬가지로 되기 때문에 중복 설명은 피한다.

[0209] 서버 장치(1) 측에 있어서, 이 경우의 CPU(11)는, 앞의 도 16에 도시한 스텝 S104의 처리 대신에 스텝 S110의 처리를, 스텝 S105의 처리 대신에 스텝 S111의 처리를 실행한다.

[0210] 스텝 S110에서 CPU(11)는, 대형 후단 네트워크에 대하여, 수신한 중간 특징 맵 IRM에 기초하여 증류 처리를 행한다. 이 증류 처리로서는, 앞의 도 18에서 설명한 바와 같은, 범용의 대형 후단 네트워크를 교사 모델로 한 증류 처리, 또는, 도 19에서 설명한 바와 같은 복수의 교사 모델 후보로서의 대형 후단 네트워크로부터 선택한 대형 후단 네트워크를 교사 모델로 한 증류 처리 중 어느 것으로 하는 것을 생각할 수 있다.

[0211] 스텝 S110에 이은 스텝 S111에서 CPU(11)는, 증류 처리에 의해 얻은 소형 후단 네트워크의 설정 데이터를 이미지 센서(30)에 송신하는 처리를 행한다.

[0212] <6. 시스템 운용 예>

[0213] 도 21을 참조하여, 정보 처리 시스템(100)의 운용예에 대하여 설명한다.

[0214] 여기에서는, 중간 특징 맵 IRM에 기초하여 선택된 후단 네트워크를 포함하는 전체 AI 모델이 에지 단말기(본 예에서는 이미지 센서(30))에 디플로이되는 예를 전제로 한다. 도 21에서는, 이 디플로이가 행해질 때까지의 운용의 스텝을 Step1부터 Step5로 나누어 설명을 행한다.

[0215] 「초기 학습」으로서 나타내는 Step1은, 초기 AI 모델(51)의 학습이다. 예를 들어, A국의 데이터베이스(DB)#0에 기억된 학습용 데이터 세트를 사용하여, B국의 운영자 클라우드에서 초기 AI 모델(51)의 학습을 행한다. 이때, 학습에 사용하는 화상 데이터에 대하여, 국가간의 월경(A국→B국)이 발생한다.

[0216] Step1에 대하여, 운영자 클라우드에 대한 초기 AI 모델(51)의 학습의 실행 지시는, C국의 AI 벤더가 행하는 것을 생각할 수 있다.

[0217] 「후단 학습」으로서 나타내는 Step2는, 커스터마이즈된 복수의 후단 네트워크를 준비하기 위한 스텝이며, B국에 있어서의 운영자 클라우드에 있어서, Step1에서 얻어진 초기 AI 모델(51)의 전단 네트워크와 후단 네트워크(잠정판)를 사용하여, 후단 네트워크의 재학습으로서 각각 다른 학습용 데이터 세트를 사용한 기계 학습(후단 학습#1, #2, #3, ・・・)을 행하여, 용도별의 복수의 후단 네트워크를 준비한다. 후단 학습에 대해서는, 예를 들어 D국용의 학습으로서 D국의 데이터베이스#1에 기억된 D국용의 학습용 데이터 세트를 사용한 학습, E국용의 학습으로서 E국의 데이터베이스#2에 기억된 E국용의 학습용 데이터 세트를 사용한 학습, F국용의 학습으로서 F국의 데이터베이스#2에 기억된 F국용의 학습용 데이터 세트를 사용한 학습 등을 행한다(이때, 화상 데이터의 월경: D로부터 F→B국이 발생).

[0218] Step2에 있어서, 후단 학습의 실행 지시는, 예를 들어 C국의 AI 벤더가 행하는 것을 생각할 수 있다.

[0219] 또한, Step2의 후단 학습은, B국이 아니라 G국의 운영자 클라우드에서 행하는 것도 생각할 수 있다(즉 Step1과는 다른 국가에서 실행하는 것도 생각할 수 있다).

[0220] 「IRM 생성」으로서 나타내는 Step3은, I국(AI 이용 현장)에 있어서의 에지(본 예에서는 카메라(3))가 촬상 화상에 기초하여 중간 특징 맵 IRM을 생성하는 스텝이다. 중간 특징 맵 IRM의 생성에는 전단 네트워크가 필요하지만, 본 예에서는, 전단 네트워크는, B국의 운영자 클라우드로부터 에지에 있어서의 이미지 센서(30)에 사전에 송신되어 있어, 이미지 센서(30) 내(AI 처리부(44))에 있어서, 촬상 화상에 기초하는 중간 특징 맵 IRM의 생성이 행해진다. 생성된 중간 특징 맵 IRM은, CPU(제어부(33))로부터 포그 서버(포그 서버(4))에 송신된 후, 포그 서버로부터 I국에 있어서의 고객 클라우드에 송신되고, 고객 클라우드로부터 운영자 클라우드에 대하여 송신된다.

[0221] 이때, 화상 데이터의 월경은 발생하지 않는다(중간 특징 맵 IRM이 되어 새니타이즈되어 있기 때문에).

[0222] Step3에 있어서, 중간 특징 맵 IRM의 생성 실행 지시는, 예를 들어 I국의 AI 이용자가 행하는 것을 생각할 수 있다.

[0223] 또한, 에지에 있어서, 후술하는 바와 같이 AI 처리는, CPU나 포그 서버가 행하는 것도 생각할 수 있고, 그 경우, 중간 특징 맵 IRM 생성을 위한 전단 네트워크의 송신은, CPU나 포그 서버를 대상으로 하여 행해진다.

[0224] 또한, 고객 클라우드는 AI 이용 현장으로서의 I국과는 다른 J국에 존재하는 경우도 생각할 수 있다.

[0225] 「후단 선택」으로서 나타내는 Step4는, 중간 특징 맵 IRM에 기초하는 후단 네트워크의 선택을 의미한다. B국(또는 G국 또는 H국)에 있어서의 운영자 클라우드가, Step3에서 에지측으로부터 송신된 중간 특징 맵 IRM에 기초하여, Step2에서 준비한 복수의 후단 네트워크 중에서 대응하는 후단 네트워크를 선택한다.

[0226] 이때, 운영자 클라우드에 대한 후단 네트워크의 선택 실행 지시는, I국에 있어서의 AI 이용자가 행하는 것을 생각할 수 있다.

[0227] Step5의 본 운용에서는, 운영자 클라우드가, 선택한 후단 네트워크와 전단 네트워크를 결합한 결합 네트워크(즉 전체 AI 모델)를 에지(이미지 센서(30), 또는 CPU나 포그 서버)에 대하여 디플로이하고, 에지에 있어서 AI 모델을 사용한 추론 처리를 행한다.

[0228] 이때, 운영자 클라우드에 대한 디플로이의 실행 지시나 에지에 대한 추론 처리의 개시 지시는, I국에 있어서의 AI 이용자가 행하는 것을 생각할 수 있다.

[0229] 여기서, 상기에서는 Step2의 「후단 학습」에 있어서, 학습용 입력 데이터로서 화상 데이터를 사용한 학습을 행하는 전제로 하였지만, 도 22의 운용예에 나타내는 바와 같이, Step2의 「후단 학습」은, 화상 데이터가 아니라 중간 특징 맵 IRM을 학습용 입력 데이터로서 사용하여 행하는 것도 생각할 수 있다.

[0230] 이 경우에는, D국, E국, F국에 있어서의 학습용 데이터 수집 현장에 대하여 Step1에서 얻어진 초기 AI 모델(51)의 전단 네트워크를 미리 배신해 두고, 각각의 국가에서 학습용 입력 데이터로서의 중간 특징 맵 IRM을 생성해 둔다. 그리고, 이 경우의 Step2의 「후단 학습」에서는, 초기 AI 모델(51)의 후단 네트워크(잠정판)에 대한 재학습을, 이들 각각의 국가에서 얻어진 중간 특징 맵 IRM을 포함하는 학습용 데이터 세트를 사용하여 실행하여, 복수의 커스텀 후단 네트워크를 생성한다.

[0231] <7. AI 커스터마이즈의 변형예>

[0232] 지금까지의 설명에서는, 초기 AI 모델(51)에 대하여, 후단 네트워크의 커스터마이즈를 행하는 예를 설명하였지만, 초기 AI 모델(51)로서의 전체 AI 모델을 재학습함으로써 커스터마이즈된 복수의 전체 AI 모델을 생성하고, 또한, 그것들 커스터마이즈된 전체 AI 모델마다, 후단 네트워크의 재학습을 행한다고 하는 것도 생각할 수 있다.

[0233] 도 23에, 구체예를 도시한다.

[0234] 도 23의 A는 초기 AI 모델(51)의 생성에 대하여 나타내고 있다. 여기에서는 초기 AI 모델(51)로서, 카메라(3)에 있어서의 이미지 센서(30)의 특성 변동을 흡수할 수 있는 AI 모델을 작성하는 예로 하고 있다. 구체적으로, 이 경우에 있어서의 초기 AI 모델(51)의 생성에서는, 학습용 데이터 세트에 있어서의 화상 데이터로서, 이미지 센서(30)의 특성 변동을 반영시킨 화상 데이터를 사용하는 것을 생각할 수 있다. 구체적으로는, 해당 화상 데이터의 세트로서, 예를 들어 복수의 이미지 센서(30)가 촬상한 화상 데이터가 혼재되는 세트를 준비하는 것을 생각할 수 있다.

[0235] 이에 의해, 이미지 센서(30)의 특성 변동에 기인하여 추론 정밀도가 변동되어 버리는 것의 방지를 도모할 수 있다.

[0236] 또한, 추론 정밀도의 변동을 흡수하기 위한 학습용 데이터 세트로서, 이미지 센서(30)의 센서 특성 이외의 변동이 있는 데이터 세트를 사용해도 된다. 예를 들어, 카메라 설치 조건이나 환경 조건에 관하여 변동이 있는 데이터 세트를 사용하는 것 등을 생각할 수 있다.

[0237] 당해 변형예에 있어서는, 상기와 같이 생성한 초기 AI 모델(51)에 대하여, 도 23의 B에 도시한 바와 같이 전체 커스텀용의 학습용 데이터 세트를 사용한 재학습을 행한다. 구체적으로는, 커스텀용의 다른 학습용 데이터 세트(도면 중에서는 #A와 #B의 2종으로 하고 있음)를 사용한 재학습을 행하여, 각각 다른 용도용으로 커스터마이즈된 전체 AI 모델로서의, 커스텀 전체 모델을 생성한다. 도면 중에서는, 학습용 데이터 세트#A를 사용한 재학습에 의해 커스텀 전체 모델(51-1)이, 학습용 데이터 세트#B를 사용한 재학습에 의해 커스텀 전체 모델(51-2)이 각각 생성된 예를 나타내고 있다.

[0238] 또한, 당해 변형예에서는, 생성된 커스텀 전체 모델마다, 그 후단 네트워크에 대하여, 커스텀용의 학습용 데이터 세트를 사용한 재학습을 행하여, 복수의 후단 네트워크를 생성한다(도 23의 C). 도면 중에서는, 커스텀 전체 모델(51-1)의 후단 네트워크에 대한 커스텀용의 학습용 데이터 세트로서 학습용 데이터 세트#a, #b, ・・・이 준비되고, 커스텀 전체 모델(51-2)의 후단 네트워크에 대한 커스텀용의 학습용 데이터 세트로서 학습용 데이터 세트#α, #β, ・・・이 준비되고, 이들의 학습용 데이터 세트를 사용한 후단 네트워크의 재학습을 행함으로써, 커스텀 전체 모델(51-1)의 후단 네트워크에 대해서는 커스텀 후단 네트워크#a, #b, ・・・이 생성되고, 커스텀 전체 모델(51-2)의 후단 네트워크에 대해서는 커스텀 후단 네트워크#α, #β, ・・・이 생성된 예를 나타내고 있다.

[0239] 당해 변형예에 있어서, 도 23의 B에서 설명한 전체 AI 모델의 재학습에 대해서는, 예를 들어 옥내 감시 용도, 옥외 감시 용도 등의 용도마다의 커스터마이즈를 위한 재학습을 행하는 것을 생각할 수 있고, 또한 도 23의 C에서 설명한 후단 네트워크의 재학습에 대해서는, 국가나 지역마다(피사체가 되는 인물의 인종의 차이 등)의 커스터마이즈를 위한 재학습을 행하는 것 등을 생각할 수 있다.

[0240] 여기서, 당해 변형예에 있어서, 도 23의 C에서 설명한 재학습에 의해 얻어진 후단 네트워크를, 다시 재학습하는 것도 생각할 수 있다. 그리고 이 경우, 2단계째의 재학습에 있어서는, 1단계째의 재학습과 전단 네트워크/후단 네트워크의 분단 위치 Dv를 다르게 하는 것도 생각할 수 있다.

[0241] 예를 들어, 전체 AI 모델(커스텀 전체 모델)이 [A][B][C]의 네트워크 구조로 이루어지는 것으로 한 경우에 있어서, 1단계째의 후단 재학습 시의 분단 위치 Dv를 [A]/[B][C]로 하는 것에 대해, 2단계째의 후단 재학습 시의 분단 위치 Dv를 보다 후단의 [A][B]/[C]로 하는 것을 생각할 수 있다. 이에 의해, 2단계째의 후단 재학습에 요하는 시간이나 통신 데이터양을, 1단계째의 후단 재학습 시보다 삭감할 수 있다.

[0242] 또한, 상기와 같은 단계적인 후단 네트워크 재학습의 예로서는, 1단계째에서는 예를 들어 미국/유럽/일본 등이 대략적인 지역마다의 커스터마이즈를 행하고, 2단계째에서는, 예를 들어 미국에 대하여, 주 단위 등의 보다 소구분의 단위로의 커스터마이즈를 행하는 것 등을 생각할 수 있다.

[0243] <8. 변형예>

[0244] 이상, 본 기술에 관한 실시 형태에 대하여 설명하였지만, 실시 형태로서는 상기한 구체예에 한정되는 것은 아니고, 다양한 변형예로서의 구성을 채용할 수 있다.

[0245] 예를 들어, 상기에서는, 후단 네트워크를 사용한 추론 처리를 이미지 센서(30) 내부에서 행하는 경우에 대응하여, 송신 처리부 F31이나 수신 처리부 F32로서의 처리를 센서 내 제어부(43)가 실행하는 예를 들었지만, 후단 네트워크를 사용한 추론 처리는, 제어부(33)로서의 프로세서(카메라(3) 내에 있어서의 이미지 센서(30) 외의 프로세서)에서 행하는 구성도 생각할 수 있고, 그 경우, 송신 처리부 F31이나 수신 처리부 F32로서의 처리는 제어부(33)로서의 프로세서가 행하게 된다.

[0246] 혹은, 후단 네트워크를 사용한 추론 처리를 포그 서버(4)의 프로세서가 행하는 구성도 생각할 수 있고, 그 경우, 송신 처리부 F31이나 수신 처리부 F32로서의 처리는 포그 서버(4)의 프로세서가 행하게 된다.

[0247] 또한, 상기에서는, AI 모델의 입력 데이터가 RGB 화상으로 되는 예를 들었지만, 본 기술에 있어서, AI 모델의 입력 데이터에는 촬상 센서에 의한 촬상 화상을 넓게 사용할 수 있다. 여기서, 본 명세서에 있어서 「촬상」이란, 피사체를 포착한 화상 데이터를 얻는 것을 넓게 의미한다. 여기에서 말하는 화상 데이터는, 복수의 화소 데이터로 이루어지는 데이터를 총칭한 것이며, 화소 데이터로서는, 피사체로부터의 수광량의 강도를 나타내는 데이터뿐만 아니라, 예를 들어 피사체까지 거리나 편광 정보, 온도의 정보 등을 넓게 포함하는 개념이다. 즉, 「촬상 센서」에 의해 얻어지는 「촬상 화상」에는, 화소마다 수광량의 강도의 정보를 나타내는 계조 화상으로서의 데이터나, 화소마다 피사체까지의 거리의 정보를 나타내는 거리 화상으로서의 데이터, 혹은 화소마다 입사광의 편광 정보를 나타내는 편광 화상으로서의 데이터, 화소마다 온도의 정보를 나타내는 서멀 화상으로서의 데이터 등을 포함하는 것이다.

[0248] 또한, AI 모델의 입력 데이터로서는, 촬상 센서에 의한 촬상 화상에 한하지 않고, 예를 들어 마이크로폰에 의해 수음된 소리 데이터 등, 촬상 화상 이외의 데이터를 사용하는 것도 생각할 수 있다.

[0249] 본 기술은, AI 모델의 입력 데이터로서, 개인 특정이 가능한 정보 내용을 갖는 데이터가 사용되는 경우에 적합해진다.

[0250] 또한, 상기에서는, 전단 네트워크의 구성을 불변으로 하는 전제로 하였지만, 전단 네트워크를 복수의 후보로부터 선택할 수 있도록 해도 된다.

[0251] 예를 들어, 촬상 센서의 종류(예를 들어, RGB 센서, IR(적외) 센서의 구별 등)에 따라서 전단 네트워크의 구성을 변화시켜야 하는 경우를 생각할 수 있고, 또한, 화상에 찍히는 인종의 종류 등에 따라서 전단 네트워크의 구성을 변화시켜야 하는 경우를 생각할 수 있다. 따라서, 이들 조건에 적합한 전단 네트워크의 선택을 행함으로써, 추론 성능의 향상을 도모할 수 있다.

[0252] 또한, 초기 AI 모델(51)을 생성하기 위한 기계 학습, 및 후보가 되는 복수의 후단 네트워크의 생성을 위한 기계 학습은, 동일 국내에서 행해지는 경우도 있고, 각각이 다른 국가에 있어서 행해지는 경우도 생각할 수 있다. 또한, 후보가 되는 복수의 후단 네트워크 생성을 위한 기계 학습은, 다른 국가에서 분담하여 행해지는 것도 생각할 수 있다.

[0253] 또한, 도 24에 도시한 바와 같이, 전단 네트워크의 출력으로서 얻어지는 추정 정보 IS를, 촬상 제어에 사용하는 것도 생각할 수 있다. 여기에서 말하는 추정 정보 IS란, 전단 네트워크로서의 AI 모델이 입력 화상에 기초하는 AI 처리를 행한 결과 얻어지는 피사체에 관한 추정 정보를 의미하는 것이며, 예를 들어 옥내에서의 촬상/옥외에서의 촬상이나, 청천 하에서의 촬상/우천 하에서의 촬상 등, 촬상 씬에 대한 추정 정보나, ROI(Region of Interest) 등의 대상 피사체 존재 범위의 추정 정보 등을 들 수 있다.

[0254] 전단 네트워크에서 얻어지는 이와 같은 추정 정보 IS를, 카메라(3)에 있어서의 촬상 제어부(65)에 부여하여, 예를 들어 셔터 스피드나 조리개값, ISO 감도 등의 밝기에 관한 촬상 설정이나, 포커스, 화이트 밸런스, 노이즈 리덕션 강도의 촬상 설정 등, 소정의 촬상 설정에 대한 제어를 행한다.

[0255] 또한, 추정 정보 IS에는, 전단 네트워크로부터 후단 네트워크의 Input Tensor로서 출력되는 중간 특징 맵 IRM이 사용되는 경우도 있을 수 있다. 예를 들어, 해당 중간 특징 맵 IRM만을 추정 정보 IS로 하는 것이나, 해당 중간 특징 맵 IRM과 다른 추정 정보를 추정 정보 IS로서 촬상 설정의 제어에 피드백하는 것도 생각할 수 있다.

[0256] <9. 실시 형태의 정리>

[0257] 이상에서 설명한 바와 같이 실시 형태로서의 제1 정보 처리 장치(서버 장치(1))는, 뉴럴 네트워크를 갖고 센서 장치(이미지 센서(30))에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 중간 특징 맵에 기초하여, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 후단 네트워크 취득부(동 F11)와, 후단 네트워크 취득부가 선택 또는 생성한 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 송신 처리부(동 F12)를 구비하는 것이다.

[0258] 상기 구성에 의하면, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서, 인공 지능 모델의 네트워크 전체가 아니라 후단 네트워크에 대해서만 선택의 후보를 작성해 두면 된다. 또는, 베이스가 되는 학습 완료 네트워크로부터 예를 들어 증류 등에 의해 대응하는 후단 네트워크를 생성하면 된다. 이때, 후보가 되는 후단 네트워크의 생성에 요하는 학습 횟수는, 전단 네트워크를 포함하는 네트워크 전체의 재학습에 요하는 학습 횟수보다도 적게 하는 것이 가능하다. 또한, 베이스가 되는 학습 완료 네트워크로부터 대응하는 후단 네트워크를 생성하는 데 요하는 학습 횟수에 대해서도, 네트워크 전체의 재학습에 요하는 학습 횟수보다도 적게 하는 것이 가능하다. 필요한 학습 횟수를 적게 할 수 있음으로써, 센서 장치의 사용 환경 조건에 적합한 후단 네트워크의 생성에 요하는 시간의 단축화를 도모할 수 있다. 또한, 필요한 학습 횟수를 적게 할 수 있음으로써, 센서 장치의 사용 환경 조건에 적합한 후단 네트워크의 생성에 요하는 학습용 입력 데이터수의 삭감을 도모할 수 있다.

[0259] 또한, 상기 구성에 의하면, 센서 장치의 사용 환경 조건에 적합한 하나의 후단 네트워크를 취득함에 있어서는, 에지측으로부터 중간 특징 맵을 취득하면 된다. 중간 특징 맵은 입력 데이터가 중간층에서 처리되어 얻어진 데이터이기 때문에, 그대로는 개인 정보를 특정하는 것이 곤란한 데이터가 된다.

[0260] 상기로부터, 실시 형태의 제1 정보 처리 장치에 의하면, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간의 단축화를 도모할 수 있고, 또한, 인공 지능 모델의 이용자가 인공 지능 모델의 적합화를 위해 클라우드측에 송신해야 할 학습용 입력 데이터의 데이터양의 삭감을 도모할 수 있기 때문에, 인공 지능 모델의 적합화에 요하는 통신 데이터양의 삭감을 도모할 수 있다.

[0261] 또한, 상기와 같이 중간 특징 맵은 그대로는 개인 정보를 특정하는 것이 곤란한 데이터로 되기 때문에, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서, 개인 정보의 유출 가능성의 저감을 도모할 수 있다.

[0262] 또한, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서, 에지측에 대해서는 후단 네트워크의 설정 데이터만을 송신하면 되고, 전단 네트워크의 설정 데이터의 송신은 불필요해져, 이 점에서도, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화시키기 위해 요하는 통신 데이터양의 삭감을 도모할 수 있다.

[0263] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 중간 특징 맵은, 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터로 되어 있다.

[0264] 2층째 이후의 중간층에 있어서는, 출력 데이터로서, 1층째의 중간층의 출력 데이터보다도 개인 정보의 특정이 곤란해지는 데이터가 얻어진다고 할 수 있다.

[0265] 따라서, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0266] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 센서 장치는 촬상 센서이며, 인공 지능 모델은 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 되어 있

다.

[0267] 상기 구성에 의하면, 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하는 인공 지능 모델에 대하여, 센서 장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간 단축화, 및 통신 데이터양의 삭감, 및 개인 정보의 유출 가능성의 저감을 도모할 수 있다.

[0268] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 되어 있다.

[0269] 즉, 예를 들어 대상 피사체가 사람인 경우에 있어서의 얼굴의 화상 영역 등, 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 정도까지 차원 압축이 행해지는 중간층의 출력 데이터이다.

[0270] 이에 의해, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0271] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로 되어 있다.

[0272] 즉, 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 정도까지 차원 압축이 행해지는 중간층의 출력 데이터이다.

[0273] 이에 의해, 개인 정보의 유출 가능성의 저감 효과를 더욱 높일 수 있다.

[0274] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 중간 특징 맵은, 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 되어 있다(도 15 등 참조).

[0275] 이에 의해, 악의의 제3자가 오토 인코더를 입수할 수 있었다고 해도, 해당 오토 인코더의 디코드부를 사용하여 중간 특징 맵으로부터 개인 정보를 특정한다고 하는 것을 불가능하게 하는 것이 가능해진다.

[0276] 따라서, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0277] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 후단 네트워크 취득부는, 중간 특징 맵에 기초하여, 후단 네트워크의 복수의 후보 중에서 하나의 후단 네트워크를 선택하고 있다(도 10, 도 16 등 참조).

[0278] 이에 의해, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서는, 인공 지능 모델의 네트워크 전체가 아니라 후단 네트워크에 대해서만 선택의 후보를 작성해 두면 되어, 전단 네트워크를 포함하는 네트워크 전체의 재학습을 행하는 경우보다도 필요한 학습 횟수를 적게 하는 것이 가능해진다.

[0279] 따라서, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간의 단축화를 도모할 수 있고, 또한, 인공 지능 모델의 이용자가 인공 지능 모델의 적합화를 위해 클라우드측에 송신해야 할 학습용 입력 데이터의 데이터양의 삭감을 도모할 수 있기 때문에, 인공 지능 모델의 적합화에 요하는 통신 데이터양의 삭감을 도모할 수 있다.

[0280] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 후단 네트워크의 복수의 후보는, 액티브 러닝으로서의 기계 학습에 의해 생성되어 있다(도 17 등 참조).

[0281] 이에 의해, 후단 네트워크의 각 후보로서 소형 네트워크를 생성하는 것이 가능해진다.

[0282] 따라서, 후단 네트워크를 사용한 추론 처리를 행하는 장치가 하드웨어 리소스가 부족한 장치인 경우에 적합해진다.

[0283] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 후단 네트워크 취득부는, 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하고 있다(도 18 내지 도 20 등 참조).

[0284] 이에 의해, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서는, 베이스가 되는 학습 완료 네트워크로부터 대응하는 후단 네트워크를 생성하면 되어, 전단 네트워크를 포함하는 네트워크 전체의 재학습을 행하는 경우보다도 필요한 학습 횟수를 적게 하는 것이 가능해진다.

[0285] 따라서, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간 단축화를 도모할 수 있고, 또한, 인공 지능 모델의 이용자가 인공 지능 모델의 적합화를 위해 클라우드측에 송신해야 할 학습용 입력 데이터의 데이터양의 삭감을 도모할 수 있기 때문에, 인공 지능 모델의 적합화에 요하는 통신 데이터양의 삭감을 도모할 수 있다.

[0286] 또한, 실시 형태로서의 제1 정보 처리 장치에 있어서는, 후단 네트워크 취득부는, 중간 특징 맵에 기초하여, 베이스가 되는 학습 완료 네트워크를 교사 모델로서 사용한 지식 증류를 행함으로써 하나의 후단 네트워크를 생성하고 있다.

[0287] 상기 구성에 의하면, 베이스가 되는 학습 완료 네트워크를 교사 모델로서 사용한 지식 증류로서, 외부 장치로부터 입력한 중간 특징 맵에 기초하는 지식 증류가 행해진다.

[0288] 이와 같은 지식 증류를 행함으로써, 센서 장치의 사용 환경 조건에 적합한 소형의 후단 네트워크를 생성할 수 있어, 후단 네트워크를 사용한 추론 처리를 행하는 장치가 하드웨어 리소스가 부족한 장치인 경우에 적합해진다.

[0289] 실시 형태로서의 제1 정보 처리 방법은, 정보 처리 장치가, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 중간 특징 맵에 기초하여, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 후단 네트워크 취득 처리와, 후단 네트워크 취득 처리에서 선택 또는 생성한 후단 네트워크의 설정 데이터를 외부에 송신하는 송신 처리를 행하는 정보 처리 방법이다.

[0290] 이와 같은 제1 정보 처리 방법에 의해서도, 상기한 실시 형태에 관한 제1 정보 처리 장치와 마찬가지의 작용 및 효과를 얻을 수 있다.

[0291] 실시 형태로서의 제2 정보 처리 장치(이미지 센서(30))는, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 송신 처리부(동 F31)와, 송신 처리부가 송신한 중간 특징 맵에 기초하여, 외부 장치가, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 송신 처리부가 송신한 중간 특징 맵에 기초하여 생성한 하나의 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 처리를 행하는 수신 처리부(동 F32)와, 수신 처리부가 수신한 설정 데이터에 의해 실현되는 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부(AI 처리부(44))를 구비하는 것이다.

[0292] 상기 구성에 의하면, 센서 장치의 검출 데이터에 기초하는 추론 처리를 행하는 에지측의 정보 처리 장치로서, 외부 장치가 센서 장치의 사용 환경 조건에 적합한 후단 네트워크를 구하기 위한 방법에, 중간 특징 맵에 기초하여 복수의 후단 네트워크의 후보 중에서 하나의 후단 네트워크를 선택하는 방법, 또는, 베이스가 되는 학습 완료 네트워크와 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 방법 중 어느 것을 채용하는 것을 가능하게 하는 정보 처리 장치가 제공된다.

[0293] 따라서, 실시 형태의 제2 정보 처리 장치에 의해서도, 전술한 제1 정보 처리 장치와 마찬가지로, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간 단축화를 도모할 수 있고, 또한, 인공 지능 모델의 이용자가 인공 지능 모델의 적합화를 위해 클라우드측에 송신해야 할 학습용 입력 데이터의 데이터양의 삭감을 도모할 수 있어, 인공 지능 모델의 적합화에 요하는 통신 데이터양의 삭감을 도모할 수 있다.

[0294] 또한, 중간 특징 맵을 사용하기 때문에, 에지측의 인공 지능 모델을 센서 장치의 사용 환경 조건에 적합화함에 있어서 개인 정보의 유출 가능성의 저감을 도모할 수 있다.

[0295] 또한, 실시 형태로서의 제2 정보 처리 장치에 있어서는, 중간 특징 맵은, 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터로 되어 있다.

[0296] 2층째 이후의 중간층에 있어서는, 출력 데이터로서, 1층째의 중간층의 출력 데이터보다도 개인 정보의 특정이 곤란해지는 데이터가 얻어지는 경향이 된다.

[0297] 따라서, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0298] 또한, 실시 형태로서의 제2 정보 처리 장치에 있어서는, 센서 장치는 촬상 센서이며, 인공 지능 모델은 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 되어 있다.

[0299] 상기 구성에 의하면, 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하는 인공 지능 모델에 대하여, 센서

장치의 사용 환경 조건에 적합화함에 있어서의 소요 시간 단축화, 및 통신 데이터양의 삭감, 및 개인 정보의 유출 가능성의 저감을 도모할 수 있다.

[0300] 또한, 실시 형태로서의 제2 정보 처리 장치에 있어서는, 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 되어 있다.

[0301] 즉, 예를 들어 대상 피사체가 사람인 경우에 있어서의 얼굴의 화상 영역 등, 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 정도까지 차원 압축이 행해지는 중간층의 출력 데이터이다.

[0302] 이에 의해, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0303] 또한, 실시 형태로서의 제2 정보 처리 장치에 있어서는, 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로 되어 있다.

[0304] 즉, 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 정도까지 차원 압축이 행해지는 중간층의 출력 데이터이다.

[0305] 이에 의해, 개인 정보의 유출 가능성의 저감 효과를 더욱 높일 수 있다.

[0306] 또한, 실시 형태로서의 제2 정보 처리 장치에 있어서는, 중간 특징 맵은, 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 되어 있다.

[0307] 이에 의해, 악의의 제3자가 오토 인코더를 입수할 수 있었다고 해도, 해당 오토 인코더의 디코드부를 사용하여 중간 특징 맵으로부터 개인 정보를 특정한다고 하는 것을 불가능하게 하는 것이 가능해진다.

[0308] 따라서, 개인 정보의 유출 가능성의 저감 효과를 높일 수 있다.

[0309] 실시 형태로서의 제2 정보 처리 방법은, 정보 처리 장치가, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 송신 처리와, 송신 처리에서 송신한 중간 특징 맵에 기초하여, 외부 장치가, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 송신 처리에서 송신한 중간 특징 맵에 기초하여 생성한 하나의 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 수신 처리와, 수신 처리에서 수신한 설정 데이터에 의해 실현되는 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리를 행하는 정보 처리 방법이다.

[0310] 이와 같은 제2 정보 처리 방법에 의해서도, 상기한 실시 형태에 관한 제2 정보 처리 장치와 마찬가지의 작용 및 효과를 얻을 수 있다.

[0311] 실시 형태로서의 정보 처리 시스템(동 100)은, 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 입력 데이터를 부여하였을 때 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 제1 송신 처리부(송신 처리부 F31)와, 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 제1 송신 처리부가 송신한 중간 특징 맵에 기초하여, 인공 지능 모델에 있어서의 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 제1 송신 처리부가 송신한 중간 특징 맵에 기초하여 하나의 후단 네트워크를 생성하는 후단 네트워크 취득부(동 F11)와, 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 후단 네트워크 취득부가 선택 또는 생성한 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 제2 송신 처리부(송신 처리부 F12)와, 제1 송신 처리부를 갖는 장치에 마련되어 설정 데이터를 수신하는 처리를 행하는 수신 처리부(동 F32)와, 제1 송신 처리부를 갖는 장치에 마련되어 수신 처리부가 수신한 설정 데이터에 의해 실현되는 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부(AI 처리부(44))를 구비하는 것이다.

[0312] 이와 같은 정보 처리 시스템에 의해서도, 상기한 실시 형태의 제1, 제2 정보 처리 장치와 마찬가지의 작용 및 효과를 얻을 수 있다.

[0313] 또한, 본 명세서에 기재된 효과는 어디까지나 예시이며 한정되는 것은 아니고, 또 다른 효과가 있어도 된다.

[0314] <10. 본 기술>

[0315] 본 기술은 이하와 같은 구성을 취할 수도 있다.

[0316] (1)

[0317] 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와,

[0318] 상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 송신 처리부를 구비하는

[0319] 정보 처리 장치.

[0320] (2)

[0321] 상기 중간 특징 맵은, 상기 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터인

[0322] 상기 (1)에 기재된 정보 처리 장치.

[0323] (3)

[0324] 상기 센서 장치는 촬상 센서이며, 상기 인공 지능 모델은 상기 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 된

[0325] 상기 (1) 또는 (2)에 기재된 정보 처리 장치.

[0326] (4)

[0327] 상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 된

[0328] 상기 (3)에 기재된 정보 처리 장치.

[0329] (5)

[0330] 상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로 된

[0331] 상기 (3)에 기재된 정보 처리 장치.

[0332] (6)

[0333] 상기 중간 특징 맵은, 상기 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 된

[0334] 상기 (1) 내지 (5) 중 어느 것에 기재된 정보 처리 장치.

[0335] (7)

[0336] 상기 후단 네트워크 취득부는, 상기 중간 특징 맵에 기초하여, 상기 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하는

[0337] 상기 (1) 내지 (6) 중 어느 것에 기재된 정보 처리 장치.

[0338] (8)

[0339] 상기 후단 네트워크의 복수의 후보는, 액티브 러닝으로서의 기계 학습에 의해 생성된

[0340] 상기 (7)에 기재된 정보 처리 장치.

[0341] (9)

[0342] 상기 후단 네트워크 취득부는, 상기 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는

[0343] 상기 (1) 내지 (6) 중 어느 것에 기재된 정보 처리 장치.

[0344] (10)

[0345] 상기 후단 네트워크 취득부는, 상기 중간 특징 맵에 기초하여, 상기 베이스가 되는 학습 완료 네트워크를 교사 모델로서 사용한 지식 증류를 행함으로써 하나의 상기 후단 네트워크를 생성하는

[0346] 상기 (9)에 기재된 정보 처리 장치.

[0347] (11)

[0348] 정보 처리 장치가,

[0349] 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부 장치로부터 입력하고, 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득 처리와,

[0350] 상기 후단 네트워크 취득 처리에서 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 송신 처리를 행하는

[0351] 정보 처리 방법.

[0352] (12)

[0353] 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 송신 처리부와,

[0354] 상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 처리를 행하는 수신 처리부와,

[0355] 상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부를 구비하는

[0356] 정보 처리 장치.

[0357] (13)

[0358] 상기 중간 특징 맵은, 상기 인공 지능 모델에 있어서의 2층째 이후의 중간층의 출력 데이터인

[0359] 상기 (12)에 기재된 정보 처리 장치.

[0360] (14)

[0361] 상기 센서 장치는 촬상 센서이며, 상기 인공 지능 모델은 상기 촬상 센서에 의해 얻어지는 촬상 화상을 입력 데이터로 하여 화상 인식 처리를 행하기 위한 인공 지능 모델로 된

[0362] 상기 (12) 또는 (13)에 기재된 정보 처리 장치.

[0363] (15)

[0364] 상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 144화소 미만이 되는 데이터로 된

[0365] 상기 (14)에 기재된 정보 처리 장치.

[0366] (16)

[0367] 상기 중간 특징 맵은, 가시화하였을 때의 개인 정보에 관한 화상 영역의 화소수가 64화소 미만이 되는 데이터로

된

[0368] 상기 (14)에 기재된 정보 처리 장치.

[0369] (17)

[0370] 상기 중간 특징 맵은, 상기 인공 지능 모델의 자기 부호화 학습에 의해 얻어지는 오토 인코더의 디코드부에 의해 디코드 불가능한 데이터로 된

[0371] 상기 (12) 내지 (16) 중 어느 것에 기재된 정보 처리 장치.

[0372] (18)

[0373] 정보 처리 장치가,

[0374] 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 송신 처리와,

[0375] 상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여, 외부 장치가, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 선택한 하나의 상기 후단 네트워크, 또는, 외부 장치가, 베이스가 되는 학습 완료 네트워크와 상기 송신 처리에서 송신한 상기 중간 특징 맵에 기초하여 생성한 하나의 상기 후단 네트워크 중 어느 후단 네트워크의 설정 데이터를 수신하는 수신 처리와,

[0376] 상기 수신 처리에서 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리를 행하는

[0377] 정보 처리 방법.

[0378] (19)

[0379] 뉴럴 네트워크를 갖고 센서 장치에 의한 검출 데이터를 입력 데이터로 하는 인공 지능 모델에 대해 상기 입력 데이터를 부여하였을 때 상기 인공 지능 모델의 소정 중간층에서 얻어지는 중간 특징 맵을 외부에 송신하는 처리를 행하는 제1 송신 처리부와,

[0380] 상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여, 상기 인공 지능 모델에 있어서의 상기 소정 중간층보다도 후단의 네트워크인 후단 네트워크의 복수의 후보 중에서 하나의 상기 후단 네트워크를 선택하거나, 또는 베이스가 되는 학습 완료 네트워크와 상기 제1 송신 처리부가 송신한 상기 중간 특징 맵에 기초하여 하나의 상기 후단 네트워크를 생성하는 후단 네트워크 취득부와,

[0381] 상기 제1 송신 처리부를 갖는 장치의 외부 장치에 마련되며, 상기 후단 네트워크 취득부가 선택 또는 생성한 상기 후단 네트워크의 설정 데이터를 외부에 송신하는 처리를 행하는 제2 송신 처리부와,

[0382] 상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 설정 데이터를 수신하는 처리를 행하는 수신 처리부와,

[0383] 상기 제1 송신 처리부를 갖는 장치에 마련되어 상기 수신 처리부가 수신한 상기 설정 데이터에 의해 실현되는 상기 후단 네트워크를 사용한 추론 처리를 행하는 추론 처리부를 구비하는

[0384] 정보 처리 시스템.

부호의 설명

[0385] 100: 정보 처리 시스템

1: 서버 장치

3: 카메라

4: 포그 서버

5: 네트워크

11: CPU

30: 이미지 센서

33: 제어부

35: 통신부

37: 버스

41: 촬상부

43: 센서 내 제어부

44: AI 처리부

47: 통신 인터페이스

48: 버스

D1, D2: 다이

50: AI 모델

50n: DNN 네트워크

51: 초기 AI 모델

55: 후 처리부

60: 오토 인코더

60a: 디코드부

Dv: 분단 위치

F11: 후단 네트워크 취득부

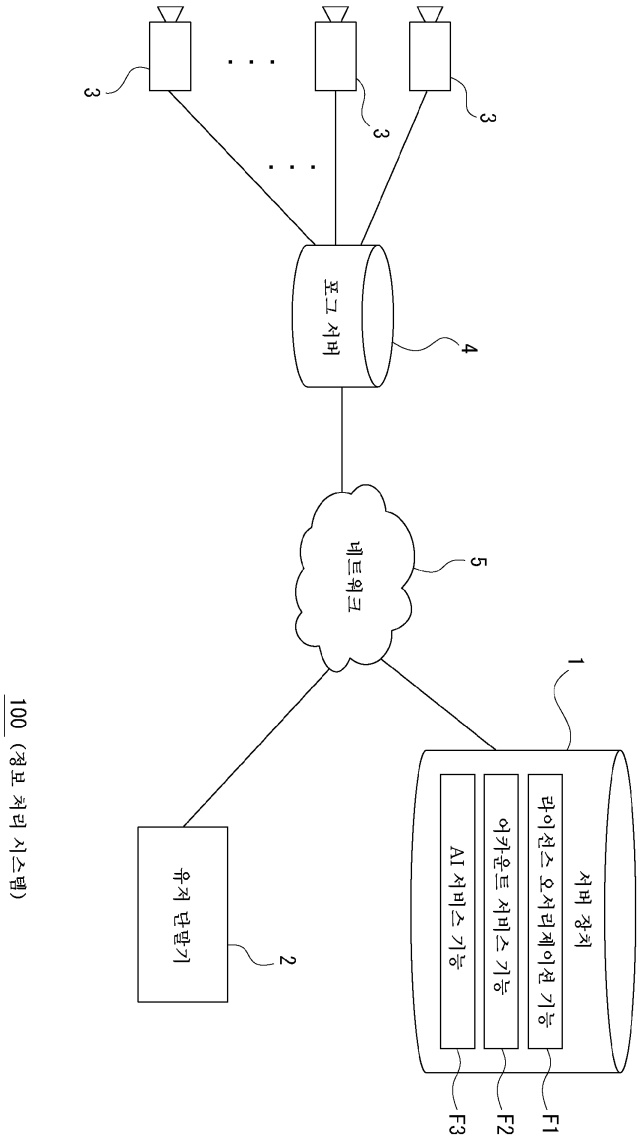
F12: 송신 처리부

F31: 송신 처리부

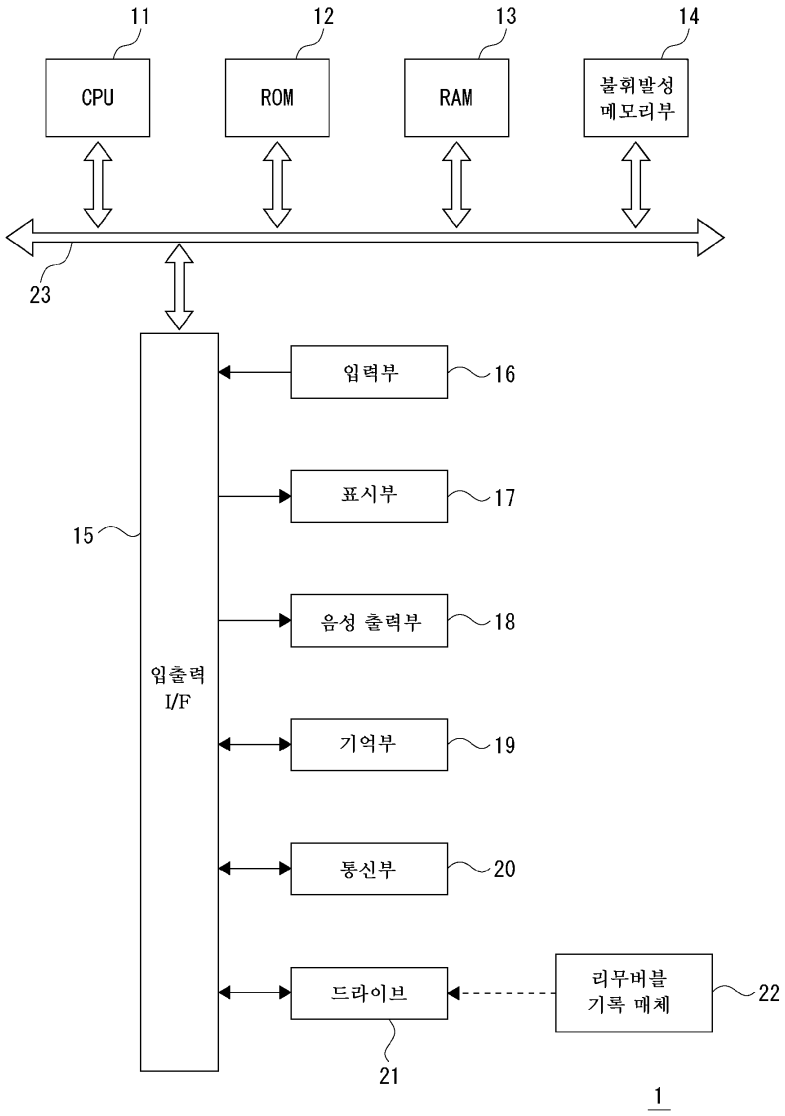
F32: 수신 처리부

도면

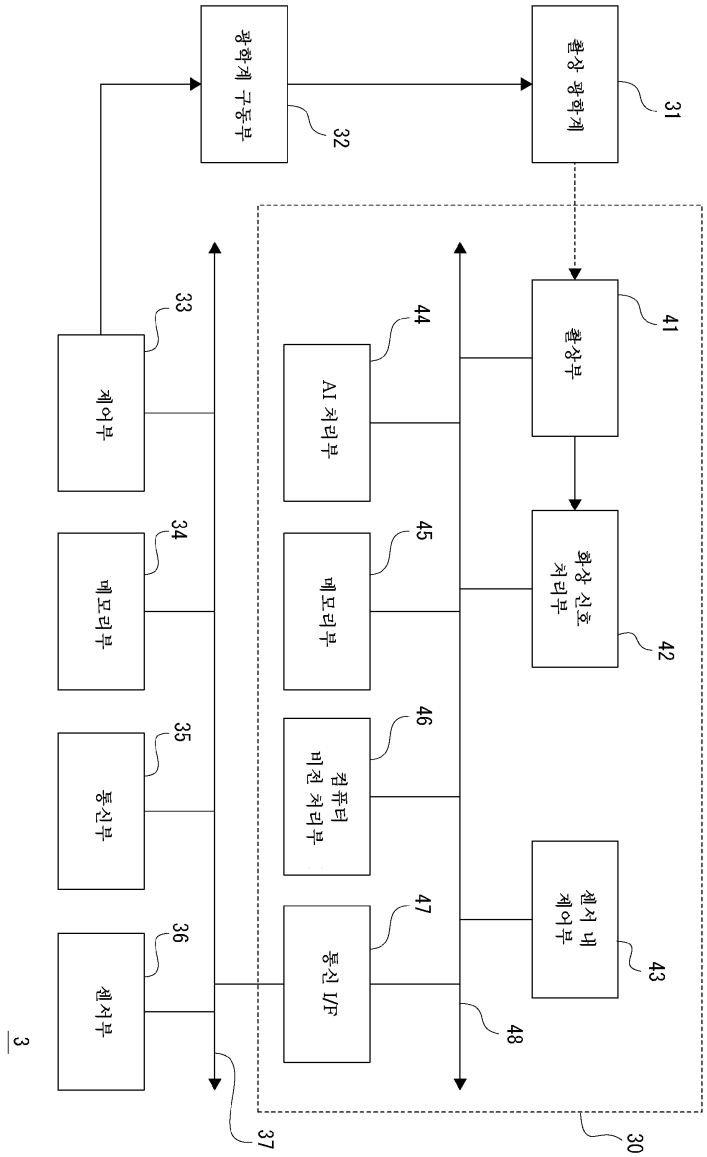
도면1


3
3
3
포그 서버
4
네트워크
5
1
서버 장치
라이선스 오서리제이션 기능
어카운트 서비스 기능
AI 서비스 기능
F1
F2
F3
유저 단말기
2
100 (정보 처리 시스템)

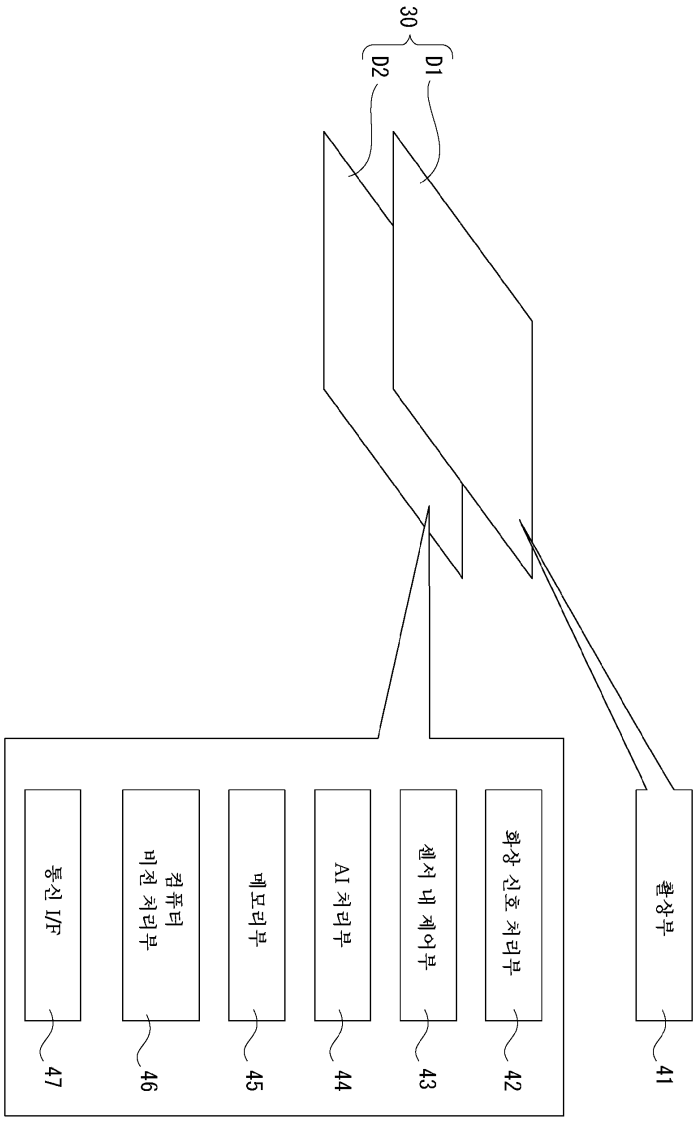
도면2


11
12
13
14
CPU
ROM
RAM
불휘발성
메모리부
23
입력부
16
표시부
17
15
음성 출력부
18
입출력
I/F
기억부
19
통신부
20
드라이브
리무버블
기록 매체
22
21
1

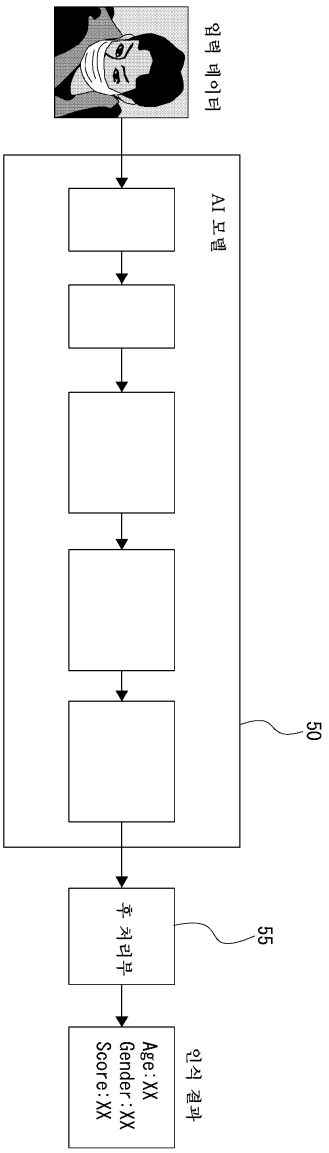
도면3


3
촬상 광학계
31
광학계 구동부
32
촬상부
41
화상 신호 처리부
42
센서 내 제어부
43
통신 I/F
47
48
컴퓨터 비전 처리부
46
메모리부
45
AI 처리부
44
제어부
33
메모리부
34
통신부
35
센서부
36
37
30

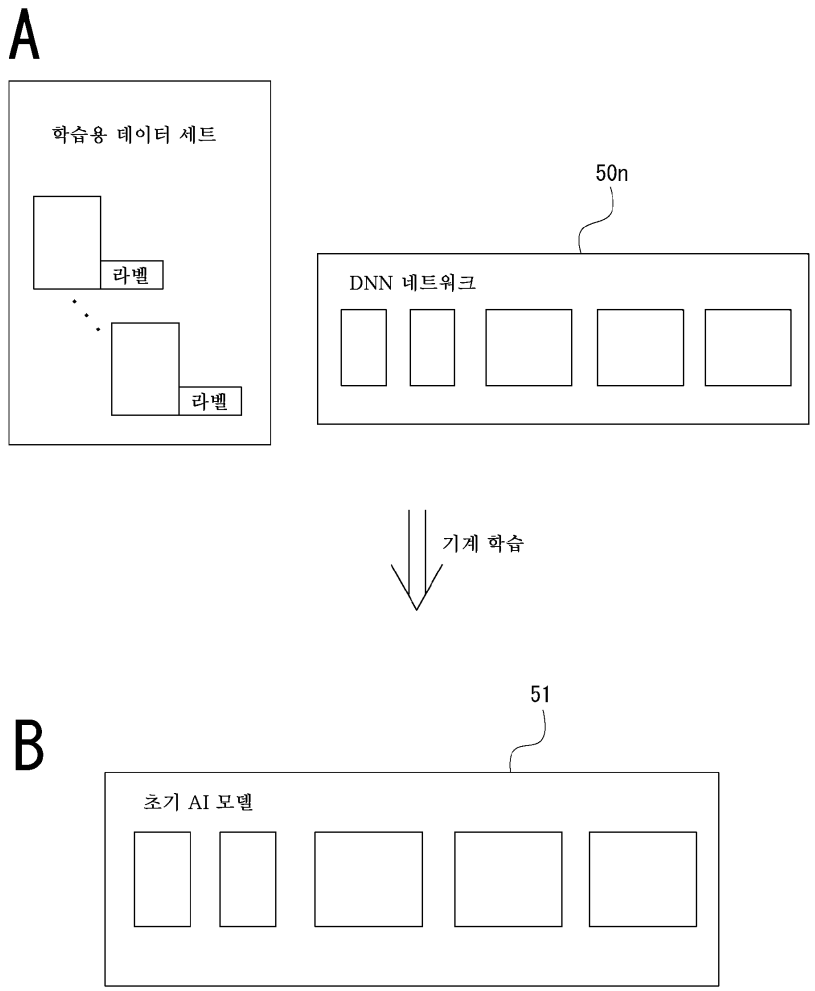
도면4


30
D1
D2
촬상부
41
화상 신호 처리부
42
센서 내 제어부
43
AI 처리부
44
메모리부
45
컴퓨터
비전 처리부
46
통신 I/F
47

도면5

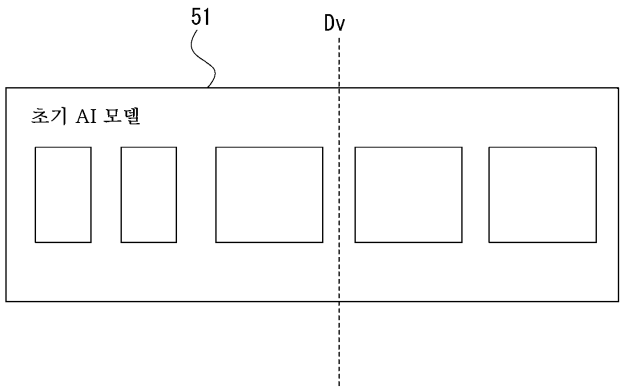


도면6

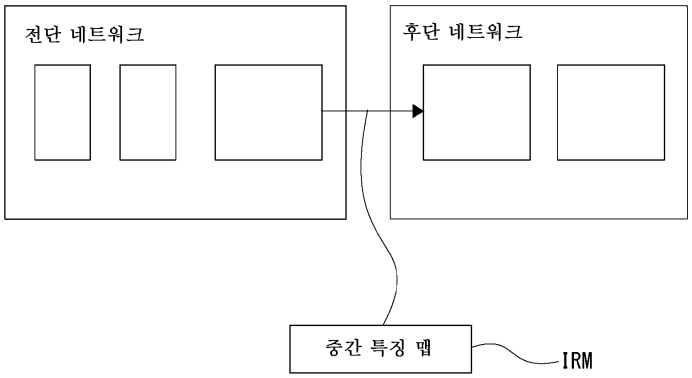


도면7

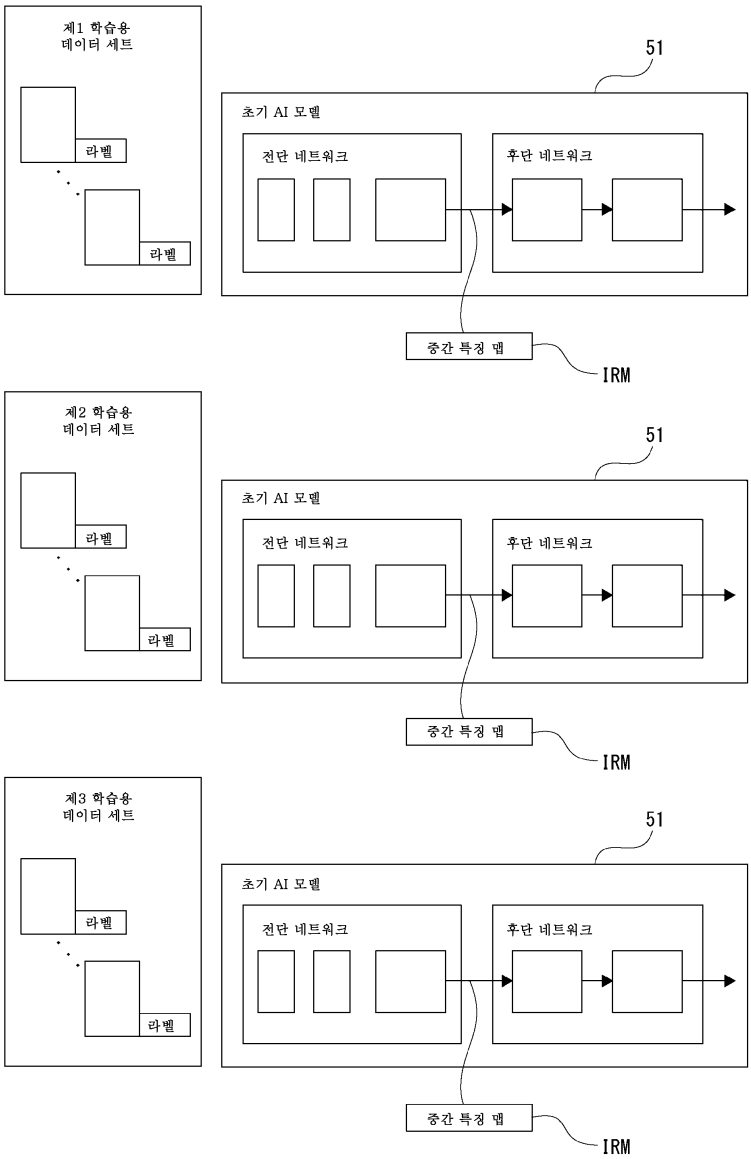
A



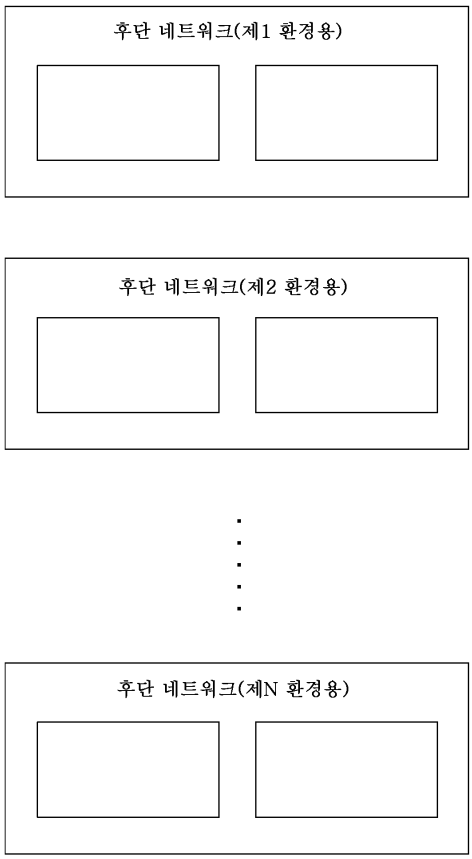
B



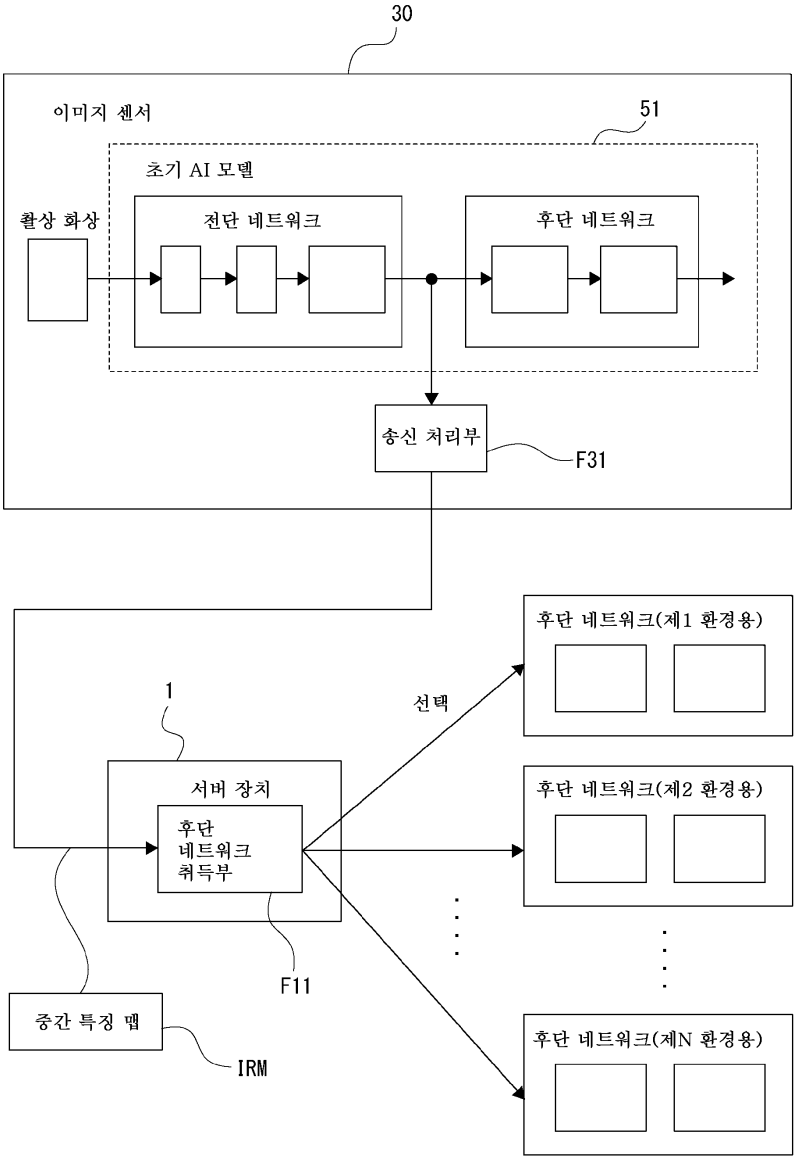
도면8


제1 학습용
데이터 세트
라벨
라벨
51
초기 AI 모델
전단 네트워크
후단 네트워크
중간 특징 맵
IRM
제2 학습용
데이터 세트
라벨
라벨
51
초기 AI 모델
전단 네트워크
후단 네트워크
중간 특징 맵
IRM
제3 학습용
데이터 세트
라벨
라벨
51
초기 AI 모델
전단 네트워크
후단 네트워크
중간 특징 맵
IRM

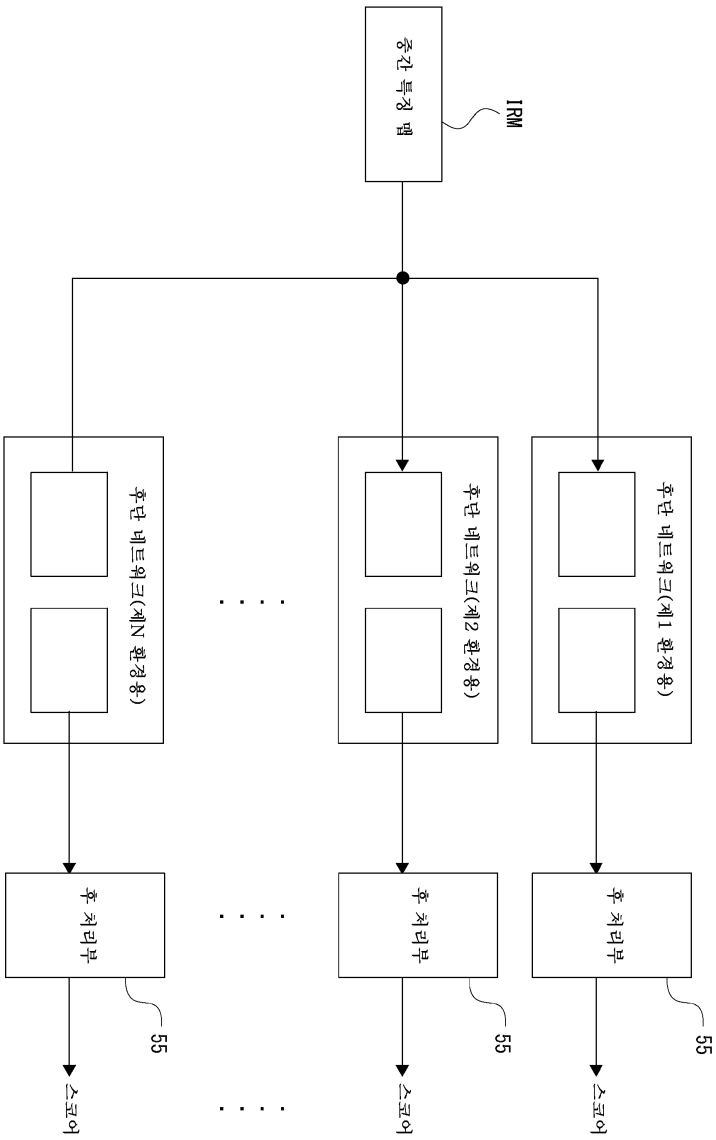
도면9



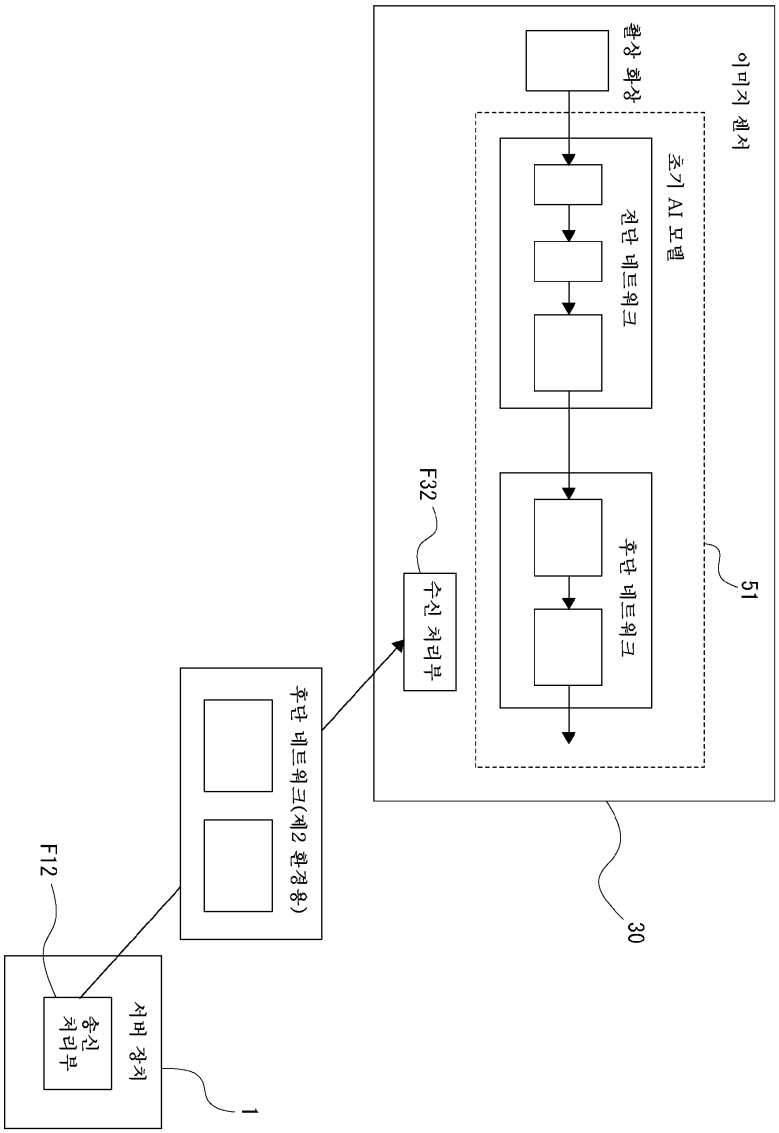
도면10


30
이미지 센서
51
초기 AI 모델
촬상 화상
전단 네트워크
후단 네트워크
송신 처리부
F31
1
서버 장치
후단
네트워크
취득부
F11
중간 특징 맵
IRM
선택
후단 네트워크(제1 환경용)
후단 네트워크(제2 환경용)
후단 네트워크(제N 환경용)

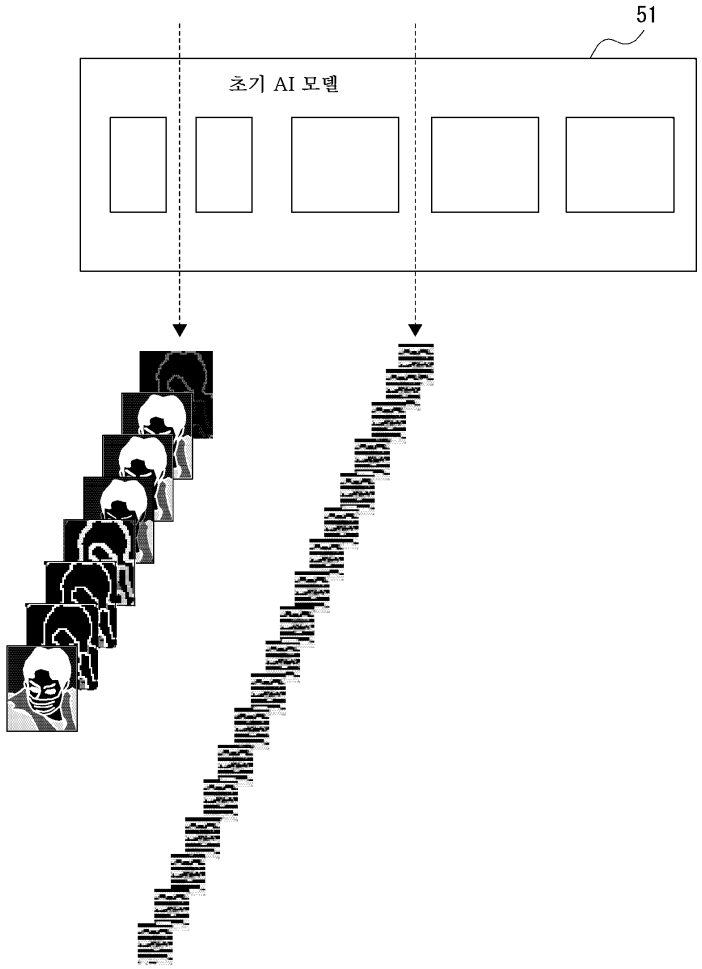
도면11


중간 특징 맵
IRM
후단 네트워크(제1 환경용)
후단 네트워크(제2 환경용)
후단 네트워크(제N 환경용)
후 처리부
후 처리부
후 처리부
55
55
55
스코어
스코어
스코어

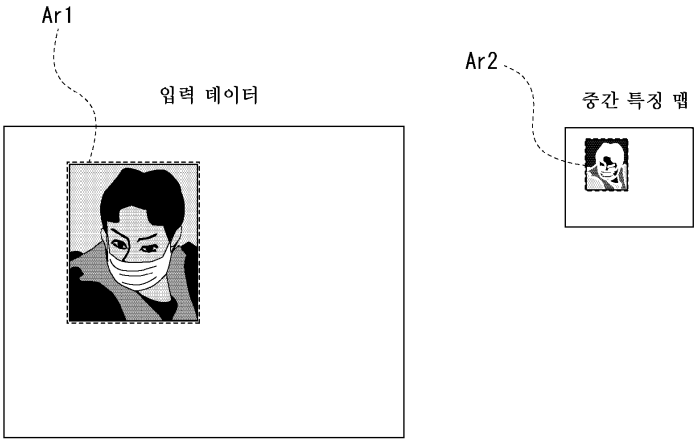
도면12


이미지 센서
촬상 화상
초기 AI 모델
전단 네트워크
후단 네트워크
51
F32
수신 처리부
30
후단 네트워크(제2 환경용)
F12
송신 처리부
처리 장치
1

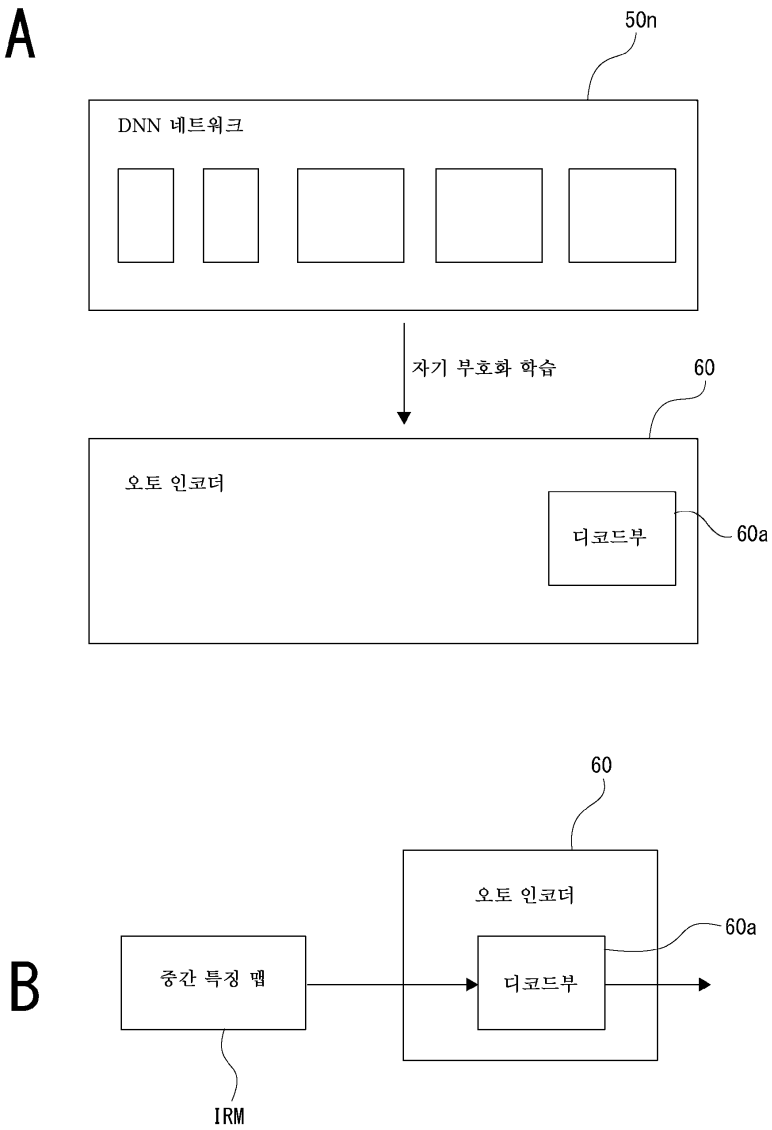
도면13



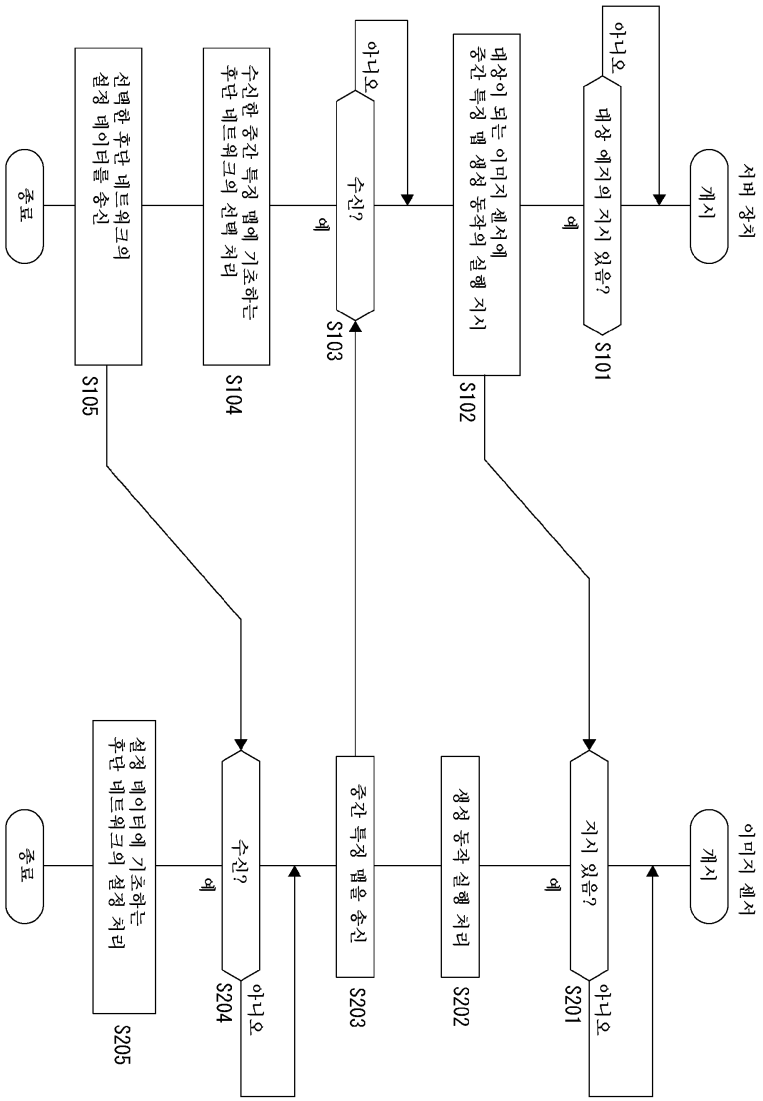
도면14



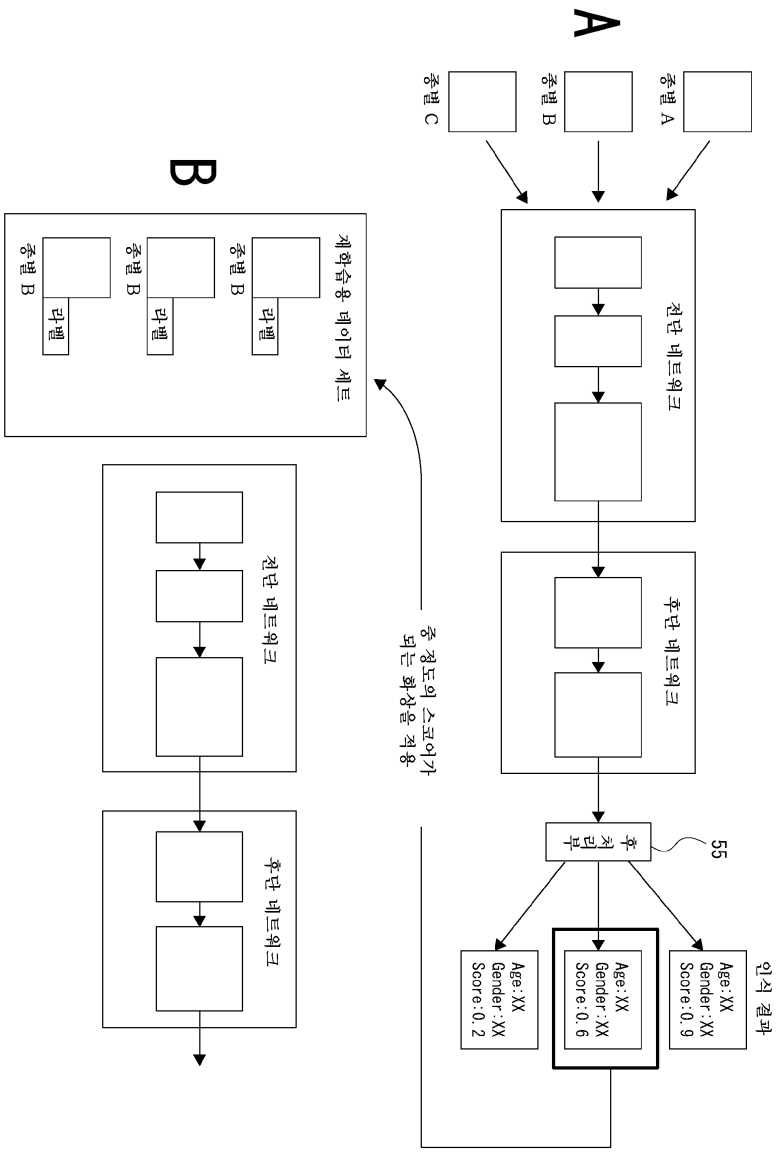
도면15


A
50n
DNN 네트워크
자기 부호화 학습
60
오토 인코더
디코드부
60a
B
60
오토 인코더
60a
중간 특징 맵
디코드부
IRM

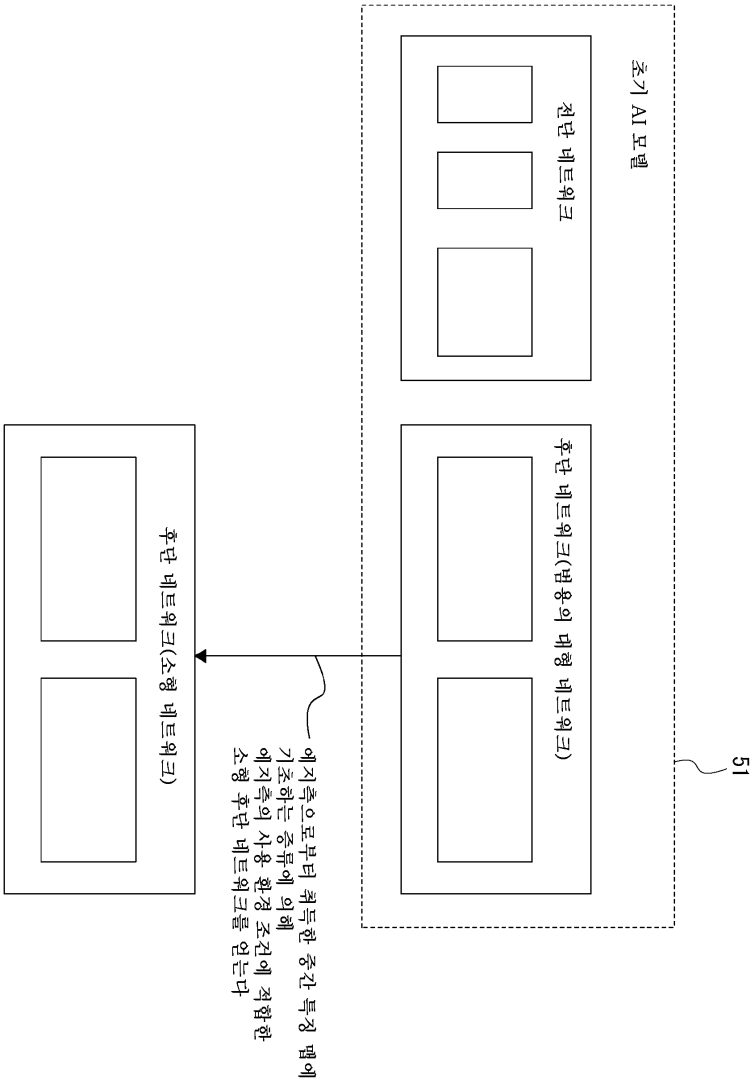
도면16


서버 장치
개시
대상 에지의 지시 있음?
S101
아니오
예
대상이 되는 이미지 센서에
중간 특징 맵 생성 동작의 실행 지시
S102
수신?
S103
아니오
예
수신한 중간 특징 맵에 기초하는
후단 네트워크의 선택 처리
S104
선택한 후단 네트워크의
설정 데이터를 송신
S105
종료
이미지 센서
개시
지시 있음?
S201
아니오
예
생성 동작 실행 처리
S202
중간 특징 맵을 송신
S203
수신?
S204
아니오
예
설정 데이터에 기초하는
후단 네트워크의 설정 처리
S205
종료

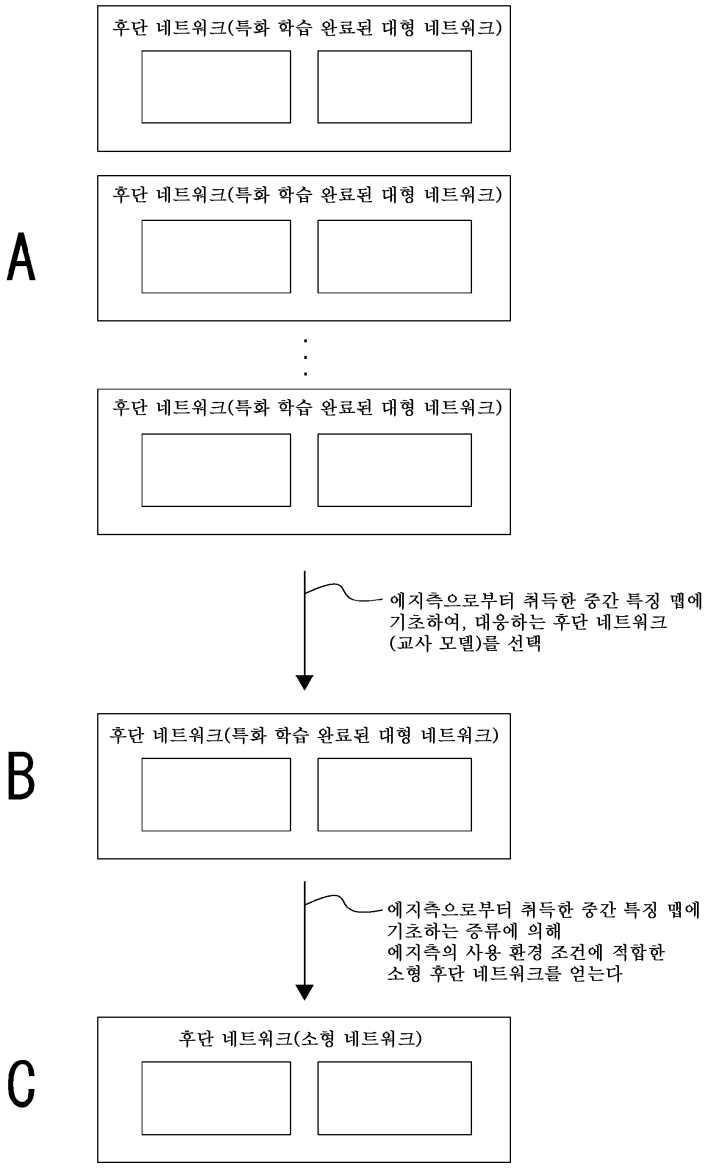
도면17


A
종별 A
종별 B
종별 C
전단 네트워크
후단 네트워크
후처리부
55
Age:XX
Gender:XX
Score:0.9
Age:XX
Gender:XX
Score:0.6
Age:XX
Gender:XX
Score:0.2
인식 결과
중 정도의 스코어가 되는 화상을 적용
B
계학습용 데이터 세트
종별 B
라벨
종별 B
라벨
종별 B
라벨
전단 네트워크
후단 네트워크

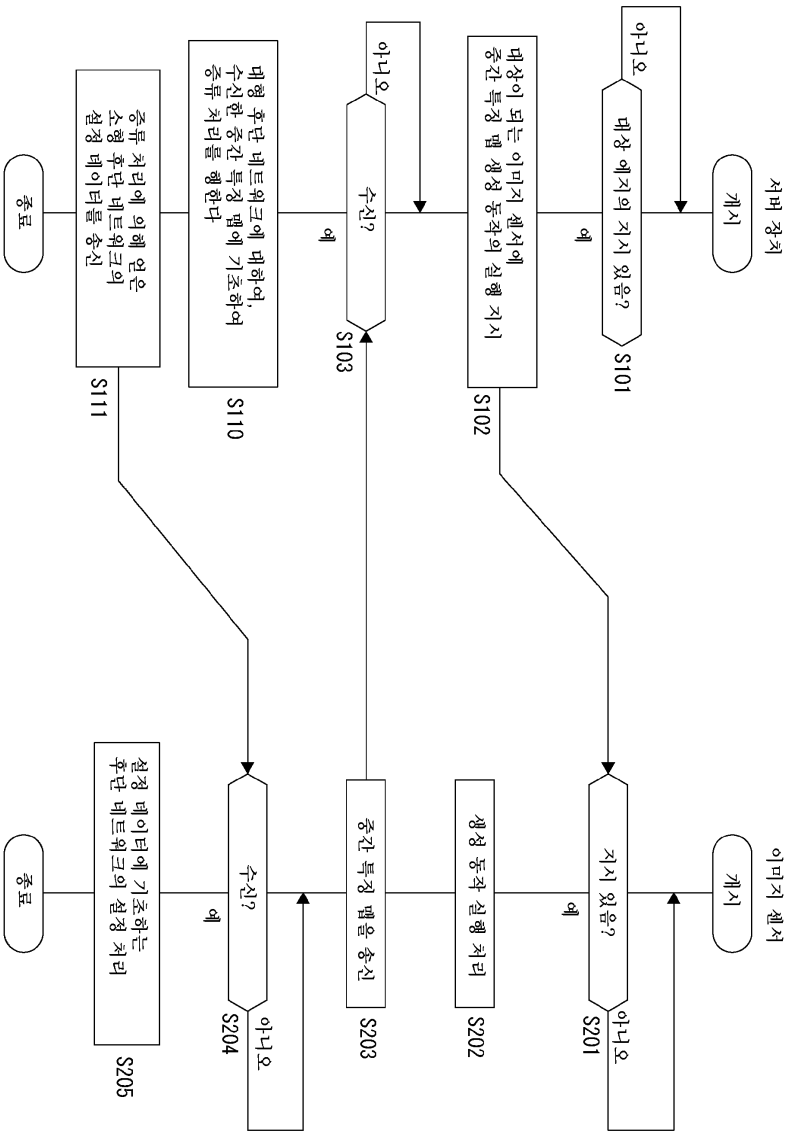
도면18


초기 AI 모델
전단 네트워크
후단 네트워크(범용의 대형 네트워크)
51
에지측으로부터 취득한 중간 특징 맵에 기초하는 승류에 의해
에지측의 사용 환경 조건에 적합한
소형 후단 네트워크를 얻는다
후단 네트워크(소형 네트워크)

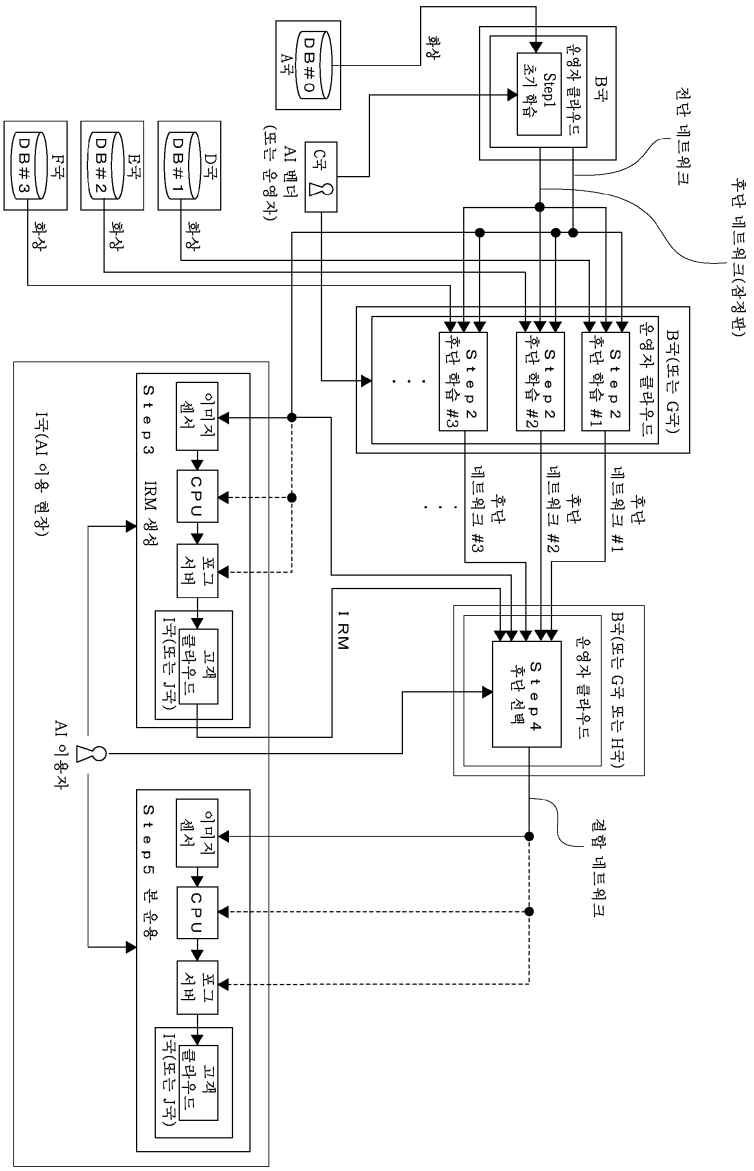
도면19


후단 네트워크(특화 학습 완료된 대형 네트워크)
후단 네트워크(특화 학습 완료된 대형 네트워크)
A
후단 네트워크(특화 학습 완료된 대형 네트워크)
에지측으로부터 취득한 중간 특징 맵에
기초하여, 대응하는 후단 네트워크
(교사 모델)를 선택
후단 네트워크(특화 학습 완료된 대형 네트워크)
B
에지측으로부터 취득한 중간 특징 맵에
기초하는 증류에 의해
에지측의 사용 환경 조건에 적합한
소형 후단 네트워크를 얻는다
후단 네트워크(소형 네트워크)
C

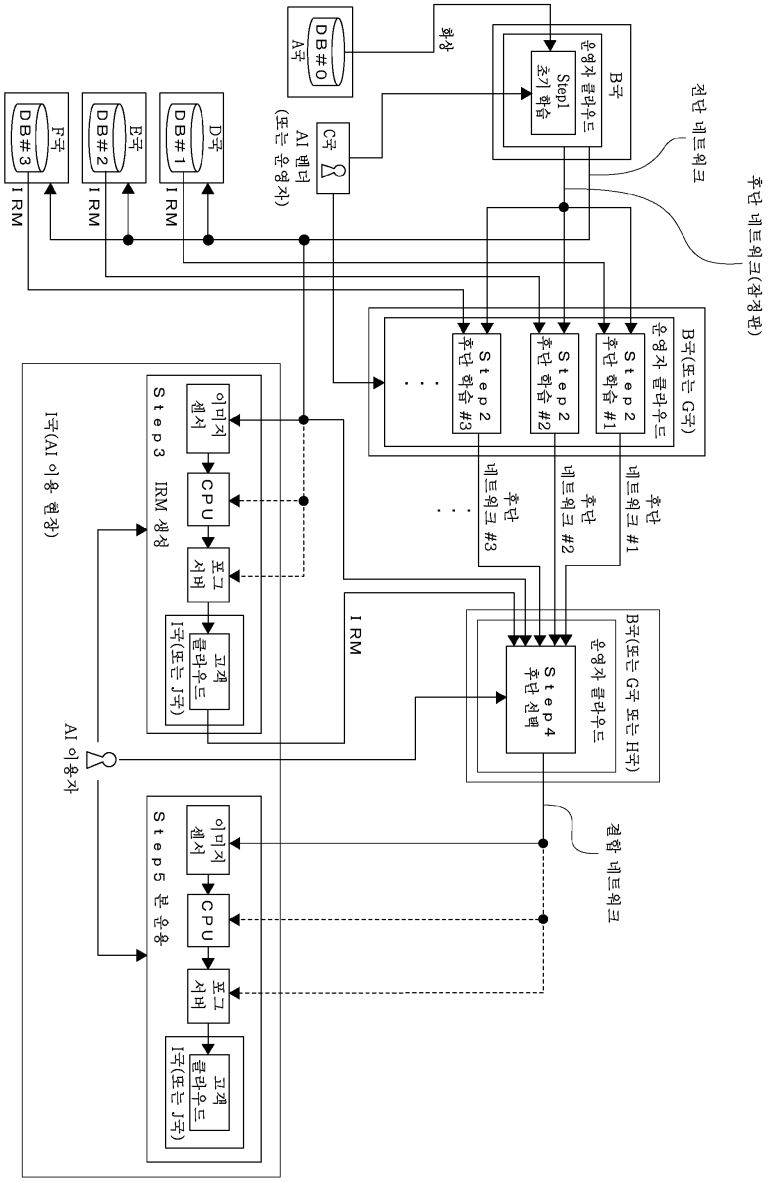
도면20


서버 장치
개시
대상 에지의 지시 있음?
S101
아니오
예
대상이 되는 이미지 센서에
중간 특징 맵 생성 동작의 실행 지시
S102
수신?
S103
아니오
예
대형 후단 네트워크에 대하여,
수신한 중간 특징 맵에 기초하여
중류 처리를 행한다
S110
중류 처리에 의해 얻은
소형 후단 네트워크의
설정 데이터를 송신
S111
종료
이미지 센서
개시
지시 있음?
S201
아니오
예
생성 동작 실행 처리
S202
중간 특징 맵을 송신
S203
수신?
S204
아니오
예
설정 데이터에 기초하는
후단 네트워크의 설정 처리
S205
종료

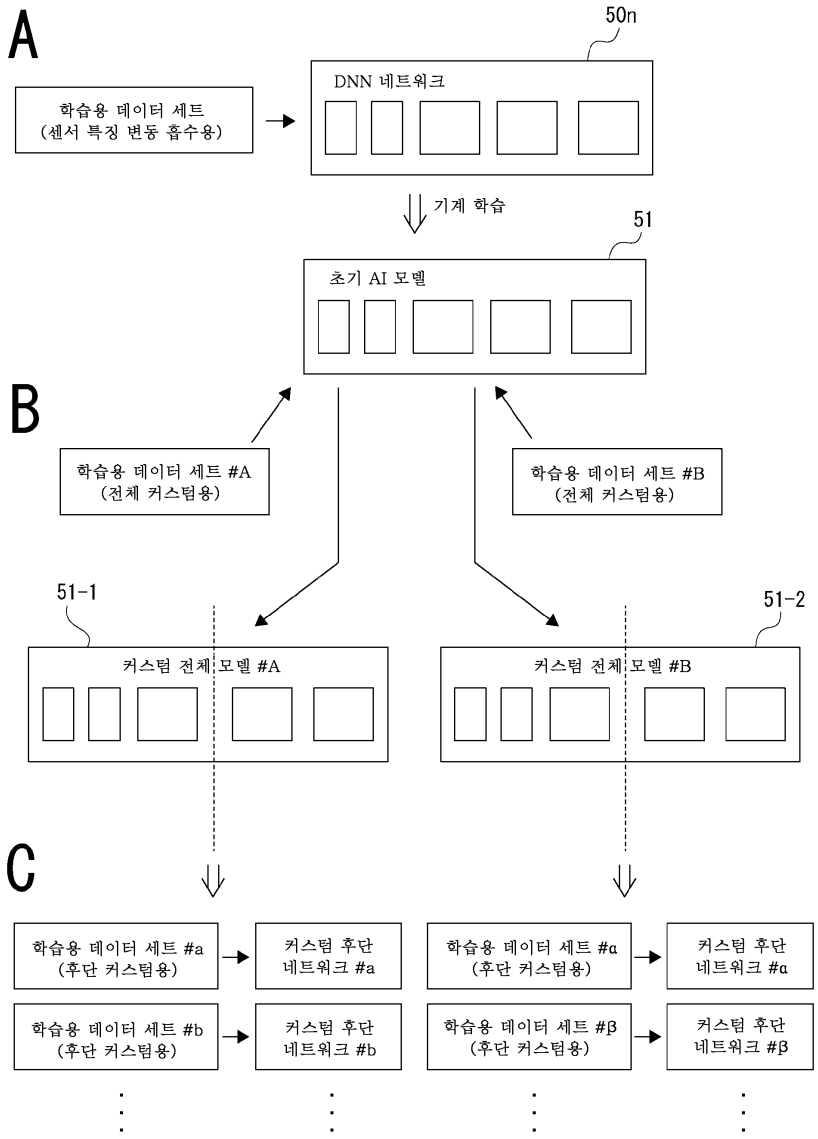
도면21


DB#0
A국
확장
B국
운영자 클라우드
Step1
초기 학습
전단 네트워크
후단 네트워크(점정판)
F국
DB#3
E국
DB#2
D국
DB#1
확장
확장
확장
AI 벤너
(또는 운영자)
C국
B국(또는 G국)
운영자 클라우드
Step2
후단 학습 #1
Step2
후단 학습 #2
Step2
후단 학습 #3
후단
네트워크 #1
후단
네트워크 #2
후단
네트워크 #3
I RM
B국(또는 G국 또는 H국)
운영자 클라우드
Step4
후단 선택
결합 네트워크
I국(AI 이용 편창)
Step3 IRM 생성
이미지
센서
CPU
포그
서버
고객
클라우드
I국(또는 J국)
AI 이용자
Step5 본 운용
이미지
센서
CPU
포그
서버
고객
클라우드
I국(또는 J국)

도면22


DB#0
A국
화상
B국
운영자 클라우드
Step1
초기 학습
전단 네트워크
후단 네트워크(참정판)
F국
DB#3
E국
DB#2
D국
DB#1
I RM
AI 벤더
(또는 운영자)
C국
B국(또는 G국)
운영자 클라우드
Step 2
후단 학습 #1
Step 2
후단 학습 #2
Step 2
후단 학습 #3
후단
네트워크 #1
후단
네트워크 #2
후단
네트워크 #3
I RM
B국(또는 G국 또는 H국)
운영자 클라우드
Step 4
후단 선택
결합 네트워크
I국(AI 이용 협정)
Step 3 IRM 생성
이미지
센서
CPU
포그
서버
고객
클라우드
I국(또는 J국)
AI 이용자
Step 5 분 운용
이미지
센서
CPU
포그
서버
고객
클라우드
I국(또는 J국)

도면23


A
50n
DNN 네트워크
학습용 데이터 세트
(센서 특징 변동 흡수용)
기계 학습
51
초기 AI 모델
B
학습용 데이터 세트 #A
(전체 커스텀용)
학습용 데이터 세트 #B
(전체 커스텀용)
51-1
51-2
커스텀 전체 모델 #A
커스텀 전체 모델 #B
C
학습용 데이터 세트 #a
(후단 커스텀용)
커스텀 후단
네트워크 #a
학습용 데이터 세트 #α
(후단 커스텀용)
커스텀 후단
네트워크 #α
학습용 데이터 세트 #b
(후단 커스텀용)
커스텀 후단
네트워크 #b
학습용 데이터 세트 #β
(후단 커스텀용)
커스텀 후단
네트워크 #β

도면24

