(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2024年12月5日(05.12.2024)

WIPO | PCT

(10) 国際公開番号
WO 2024/247193 A1

(51) 国際特許分類:
G06T 7/00 (2017.01) *G06T 7/20* (2017.01)

(21) 国際出願番号： PCT/JP2023/020354

(22) 国際出願日： 2023年5月31日(31.05.2023)

(25) 国際出願の言語： 日本語

(26) 国際公開の言語： 日本語

(71) 出願人: 日本電信電話株式会社 (**NIPPON TELEGRAPH AND TELEPHONE CORPORATION**) [JP/JP]; 〒1008116 東京都千代田区大手町一丁目５番１号 Tokyo (JP).

(72) 発明者: 永井 隆昌 (**NAGAI, Takasuke**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 武田 翔一郎 (**TAKEDA, Shoichiro**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 江崎 健司 (**ESAKI, Kenji**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP). 瀬下 仁志 (**SESHIMO, Hitoshi**); 〒1808585 東京都武蔵野市緑町３丁目９－１１ ＮＴＴ 知的財産センタ内 Tokyo (JP).

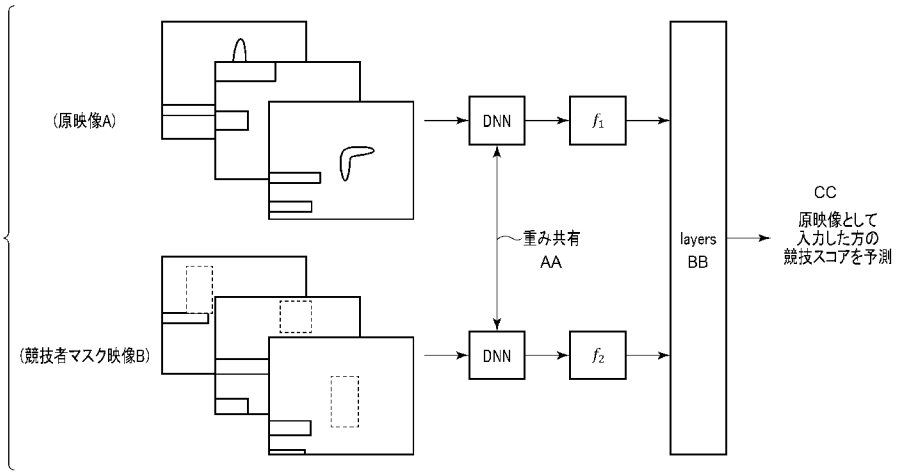
(74) 代理人: 弁理士法人鈴榮特許綜合事務所 (**SUZUYE & SUZUYE**); 〒1050014 東京都港区芝三丁目２３番１号 セレスティン芝三井ビルディング１１階 Tokyo (JP).

(81) 指定国(表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,

(54) **Title:** IMAGE EVALUATION DEVICE, METHOD, AND PROGRAM

(54) 発明の名称： 画像評価装置、方法およびプログラム

［図6］



A Original video
B Competitor mask video
AA Share weights
BB layers
CC Predict competition score of original video as input

(57) **Abstract:** An image evaluation device according to one embodiment of the present invention comprises: a first update unit that inputs a moving image related to an action of a first person and a moving image related to an action of a second person to a model for determining a feature of each moving image, calculates an estimated value of a difference between an action evaluation score of the moving image related to the action of the first person and an action evaluation score of the moving image related to the action of the second person on the basis of each feature output from the model, and updates

WO 2024/247193 A1

［続葉有］

BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 指定国(表示のない限り、全ての種類の広域保護が可能):ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

添付公開書類:

一　国際調査報告（条約第21条(3)）

parameters of the model so that the estimated value approaches a true value; and a second update unit that inputs, to the model, moving images resulting from the moving image related to the action of the first person and/or the moving image related to the action of the second person being subjected to mask processing, calculates an estimated value of a difference between evaluation scores of the actions on the basis of each feature output from the model, and updates parameters of the model so that the estimated value approaches a true value.

(57) 要約：一実施形態に係る画像評価装置は、第１の人物の動作に係る動画像および第２の人物の動作に係る動画像を、それぞれの動画像の特徴を求めるモデルに入力し、モデルから出力された特徴に基づいて、第１の人物の動作に係る動画像での動作の評価スコアおよび第２の人物の動作に係る動画像での動作の評価スコアの差分の推定値を計算し、この推定値が真値に近づくように、モデルのパラメータを更新する第１の更新部と、第１の人物の動作に係る動画像および第２の人物の動作に係る動画像の少なくとも一方に対してマスク処理が施されたときのそれぞれの動画像をモデルに入力し、このモデルから出力された特徴に基づいて、係る動作の評価スコアの差分の推定値を計算し、この推定値が真値に近づくように、モデルのパラメータを更新する第２の更新部と、を有する。

明　細　書

発明の名称 ： 画像評価装置、方法およびプログラム

技術分野

[0001] 本発明の実施形態は、画像評価装置、方法およびプログラムに関する。

背景技術

[0002] 従来より、例えば採点競技または医学生の模擬手術などの、動画像（映像）中のヒトの動作品質を自動で評価する手法としてのＡＱＡ（Action Quality Assessment）がある。

この手法では、例えば上記の動作品質の良し悪しを示すスコアをＤＮＮ（Deep Neural Network）などを用いたモデルにより予測する。また、この予測したスコアと真値スコアの差分に応じて、この真値スコアを予測できるように上記モデルのパラメータを学習する。

[0003] また、例えば非特許文献１に開示されるように、スコアが異なる二つの動作、すなわち同じ技を入力して、スコアの差を予測するような学習により、似たような動作間の細かい品質の差に特に注目するような技術がある。

[0004] この技術では、例えば、入力データをモデルに入力して推定スコアを求め、この推定スコアと真値スコアに基づく損失、ここでは推定スコアと真値スコアのＭＳＥ（Mean Squared Error：平均二乗誤差）を計算し、誤差逆伝搬法（Backpropagation）によりモデルのパラメータを更新する。

先行技術文献

非特許文献

[0005] 非特許文献1：Yu et al., ”Group-aware Contrastive Regression for Action Quality Assessment”, In ICCV. 2021.

発明の概要

発明が解決しようとする課題

[0006] しかし、上記のような従来の手法では、ヒトの動作に関する特徴、例えば関節座標などが明示的に与えられず、動画像の全体のみを学習用のデータと

して用いているため、ヒトの動作に関係ない特徴、例えば競技会場などの背景の差を抽出してしまい、ヒトの動作に汎化することができない。

[0007] 例えば、ヒトの動作を含む、あるシーンの第１の原映像と、シーンが異なる第２の原映像を、パラメータの重み付けを共有した個々のモデルに入力して、各原映像に応じたスコアの差を予測するときに、第１の原映像と第２の原映像とで背景が異なる場合に、この背景の差に注目して上記のスコアの差が予測されてしまうことがあり、ヒトの動作品質を適切に評価できない。

[0008] この発明は、上記事情に着目してなされたもので、その目的とするところは、動画像中のヒトの動作品質を適切に評価することができるようにした画像評価装置、方法およびプログラムを提供することにある。

課題を解決するための手段

[0009] 本発明の一態様に係る画像評価装置は、第１の人物の動作に係る動画像および第２の人物の動作に係る動画像を、それぞれの動画像の特徴を求めるモデルに入力し、前記モデルから出力された特徴に基づいて、前記第１の人物の動作に係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作に係る動画像での前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新する第１の更新部と、前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の少なくとも一方に対して人物に係る領域にマスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新する第２の更新部と、を備える。

[0010] 本発明の一態様に係る画像評価方法は、画像評価装置により行なわれる方法であって、前記画像評価装置の第１の更新部により、第１の人物の動作に係る動画像および第２の人物の動作に係る動画像を、それぞれの動画像の特

徴を求めるモデルに入力し、前記モデルから出力された特徴に基づいて、前記第１の人物の動作に係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作に係る動画像での前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新することと、前記画像評価装置の第２の更新部により、前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の少なくとも一方に対して人物に係る領域にマスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新することと、を備える。

発明の効果

[0011] 本発明によれば、動画像中のヒトの動作品質を適切に評価することができる。

図面の簡単な説明

[0012] [図1]図１は、本発明の一実施形態に係る画像評価装置の適用例を示す図である。

[図2]図２は、原映像の一例を示す図である。

[図3]図３は、競技者マスク映像の一例を示す図である。

[図4]図４は、学習部による第１のパターンの学習処理の一例を示す図である。

[図5]図５は、学習部による第２のパターンの学習処理の一例を示す図である。

[図6]図６は、学習部による第３のパターンの学習処理の一例を示す図である。

[図7]図７は、本発明の一実施形態に係る画像評価装置のハードウエア（hardware）構成の一例を示すブロック図（block diagram）である。

発明を実施するための形態

[0013] 以下、図面を参照しながら、この発明に係わる一実施形態を説明する。

図１は、本発明の一実施形態に係る画像評価装置の適用例を示す図である。

図１に示されるように、本発明の一実施形態に係る画像評価装置（映像評価装置）１００は、学習処理部１０および推定処理部２０を備える。学習処理部１０は、入力部１１、学習部（更新部）１２および学習モデルＤＢ（データベース）１３を有する。推定処理部２０は、入力部２１、推定部２２および学習モデルＤＢ２３を有する。

なお、学習処理部１０の学習モデルＤＢ１３と推定処理部２０の学習モデルＤＢ２３には、ヒトの動作を含む複数の動画像の特徴の抽出に用いられるモデルが格納される。また、これらは共通した、すなわち１つのモデルＤＢであっても良い。

[0014] 推定処理部２０の入力部２１は、ヒトの動作を含む複数の動画像を入力する。推定部２２は、この入力した動画像を、学習モデルＤＢ２３に格納されるモデルに入力し、このモデルからの出力である、動画像中のヒトの動作評価のスコアを推定する。

[0015] また、推定処理部２０の入力部２１は、学習モデルＤＢ１３に格納されるモデルのパラメータの学習を目的とする、ヒトの動作を含む複数の動画像を入力し、この動画像中のヒトの動作評価のスコアの真値も入力する。学習部１２は、この入力した動画像を、学習モデルＤＢ１３に格納されるモデルに入力し、このモデルからの出力である、動画像中のヒトの動作評価のスコアを推定する。

そして、学習部１２は、この推定したスコアと上記スコアの真値との差分に基づいて、この差分がゼロに近づく、またはゼロになるように上記モデルのパラメータを更新して、この更新したモデルを推定処理部２０の学習モデルＤＢ２３に格納する。

[0016] 本実施形態では、原映像、およびこの原映像におけるヒト、ここでは競技

者の動作が映る領域がマスクされた競技者マスク映像でなる２種類の映像を用いて、モデルのパラメータの学習を行なう。

[0017] 図２は、原映像の一例を示す図である。

本実施形態では、原映像として、図２に示される原映像Ａおよび原映像Ｂが、モデルのパラメータの学習に用いられ得る。

[0018] 原映像Ａは、水泳競技での高飛び込みの競技者Ａの動作、すなわち高飛び込み台から水面に飛び込む競技者の動作が映るオリジナルの映像である。

原映像Ｂは、上記高飛び込みの競技者Ｂの動作が映るオリジナルの映像であり、原映像Ａに対し、競技者および競技のタイミングが異なる。原映像Ａと原映像Ｂでの競技者の動作は高飛び込みの同じ種別の技の実行に係る動作であるとする。

[0019] 図３は、競技者マスク映像の一例を示す図である。

本実施形態では、競技者マスク映像として、図３に示される競技者マスク映像Ａおよび競技者マスク映像Ｂが、モデルのパラメータの学習に用いられ得る。

[0020] 競技者マスク映像Ａは、上記原映像Ａにおける競技者Ａの領域が四角形の領域で塗りつぶされてマスクされることで外部または学習処理部１０内で生成された映像である。また、競技者マスク映像Ｂは、上記原映像Ｂにおける競技者Ｂの領域がマスクされることで外部または学習処理部１０内で生成された映像である。

[0021] 学習処理部１０の入力部１１は、原映像Ａ、競技者マスク映像Ａ、原映像Ｂ、競技者マスク映像Ｂ、競技スコアＡの真値、および競技スコアＢの真値を入力して、学習部１２に渡す。競技スコアＡの真値は、原映像Ａでの競技者の動作に対する評価のスコアの真値である。競技スコアＢの真値は、原映像Ｂでの競技者の動作に対する評価のスコアの真値である。

原映像ＡおよびＢにおける競技者の数は一人でも複数人でも良い。例えば原映像は、シンクロナイズドスイミング（synchronized swimming）などの、複数の競技者が映る映像でも良い。

[0022] 競技者マスク映像ＡおよびＢの生成に係る、競技者の領域を取得する方法は特に限られない。例えば「K. He et al. Mask R-CNN. In ICCV, 2017.」に開示された技術を用いて自動で生成しても良いし、原映像の全てのフレームを目視で確認しながら手動で生成しても良い。

また、上記の評価のスコアの与え方は特に限られない。例えば映像中の動作に対して審判により採点された点数が用いられても良い。

[0023] 以下、学習部１２による学習処理について説明する。本実施形態では、学習部１２は、以下の第１のパターンの学習処理、第２のパターンの学習処理、および第３のパターンの学習処理をそれぞれ行なうことで、複数回にわたってモデルのパラメータを学習する。

[0024] 図４は、学習部による第１のパターンの学習処理の一例を示す図である。

この第１のパターンの学習処理では、学習部１２は、従来の手法のように、原映像ＡおよびＢ、すなわち複数の原映像同士をモデルに入力し、このモデルから出力される競技スコア同士の差が適切に予測されるようにモデルのパラメータが学習される。この差分は、原映像Ａでの競技者の動作に対する評価のスコアの予測値である競技スコアＡと、原映像Ｂでの競技者の動作に対する評価のスコアの予測値である競技スコアＢとの差分である。

[0025] 図４に示されるように、本実施形態では、学習モデルＤＢ１３に格納されるモデルは、第１のＤＮＮおよび第２のＤＮＮを含む。第１のＤＮＮは１つ目の映像を入力可能で、第２のＤＮＮは２つ目の映像を入力可能である。モデルのパラメータは、上記第１のＤＮＮのノード間の重み付けと第２のＤＮＮのノード間の重み付けを含み、これらの重み付けは互いに共有される。

[0026] 図４に示された例では、第１のＤＮＮから原映像Ａの特徴である特徴f_1が出力され、第２のＤＮＮから原映像Ｂの特徴である特徴f_2が出力される。これらの特徴は学習部１２内のlayersに入力され、このlayersから競技スコアＡと競技スコアＢの差の予測値が出力される。

そして、第１のパターンの学習処理では、この予測値が、競技スコアＡの真値と競技スコアＢの真値との差に近づく、または同じとなるように、モデ

ルのパラメータが学習される。

[0027] 図５は、学習部による第２のパターンの学習処理の一例を示す図である。

この第２のパターンの学習処理では、学習部１２は、競技者マスク映像Ａおよび競技者マスク映像Ｂ、すなわち複数の競技者マスク映像同士をモデルに入力し、競技者が見えない場合にはスコアは０点であると仮定して、互いの競技スコアの差として０点が予測されるようにモデルのパラメータが学習される。

[0028] 図５に示された例では、第１のＤＮＮから競技者マスク映像Ａの特徴である特徴f_1が出力され、第２のＤＮＮから競技者マスク映像Ｂの特徴である特徴f_2が出力される。これらの特徴は学習部１２内のlayersに入力され、このlayersから競技スコアの差の予測値が出力される。

そして、第２のパターンの学習処理では、この予測値が競技スコアの真値としての０点に近づく、または０点になるように、モデルのパラメータが学習される。

[0029] 図６は、学習部による第３のパターンの学習処理の一例を示す図である。

この第３のパターンの学習処理では、学習部１２は、原映像Ａと競技者マスク映像Ｂの組み合わせ、または原映像Ｂと競技者マスク映像Ａの組み合わせを入力し、競技者が見えない場合には競技スコアは０点であると仮定して、互いの競技スコアの差として、入力した原映像の競技スコアが適切に予測されるように、モデルのパラメータが学習される。

[0030] 図６に示された例では、原映像Ａと競技者マスク映像Ｂの組み合わせ、または原映像Ｂと競技者マスク映像Ａの組み合わせが入力され、第１のＤＮＮから原映像Ａの特徴である特徴f_1が出力され、第２のＤＮＮから競技者マスク映像Ｂの特徴である特徴f_2が出力される。これらの特徴は学習部１２内のlayersに入力され、このlayersから競技スコアの差の予測値が出力される。

そして、第３のパターンの学習処理では、この予測値が、原映像Ａでの競技者の動作に対する競技スコアの真値に近づく、または同じとなるように、モデルのパラメータが学習される。

[0031] 上記の第１、第２、及び第３のパターンの学習処理のいずれにおいても、ＤＮＮおよびlayersの構造は限定されない。

また、上記第１のおよび第２のＤＮＮには、例えば「Carreira et al. “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset.”, In CVPR 2017」に開示されるＩ３Ｄ（Two-Stream Inflated 3D Convnet）が適用されても良い。

[0032] また、上記layersは複数の全結合層でも良いし、上記非特許文献１に開示されたモデル構造が適用されても良い。

[0033] 学習処理では、予測した競技スコアと競技スコアの真値との差分から損失関数を計算し、誤差逆伝搬によりネットワークモデルの学習を行なう。この学習したモデルのパラメータの重み付けは学習モデルＤＢ１３に保存され、推定処理部２０の学習モデルＤＢ２３にも保存される。損失関数の種別は特に限られない。例えばL1 distance、L2 distance、あるいは，これらの合計が損失関数として用いられても良い。

[0034] 上記の第１、第２および第３のパターンによる学習処理の実行の順番は特に限られない。また、各種パターンによる学習処理におけるパラメータの重み付けの更新のタイミングは特に限られない。

例えば、最初の50エポック（epoch）では学習を安定させるために上記第１のパターンの学習処理により学習し、続く50エポックでは上記第１、第２および第３のパターンの全てによる学習処理により学習するなど、学習のエポック数、すなわち学習回数に応じて組み合わせを変化させても良い。

[0035] また、エポックを通じて常にランダムに学習処理のパターンの種別が採用されても良い。例えば第１のエポック内で上記第１、第２および第３のパターンの全ての学習処理が行なわれても良いし、第１のエポックでは上記第１のパターン、第２のエポックでは上記第２のパターンというように、エポックごとに、いずれか一つのパターンにより学習処理が行われても良い。

[0036] 次に、上記の学習処理が実施された後の、推定処理部２０による推定処理について説明する。

この推定処理では、原映像Ａと原映像Ｂとの組み合わせが入力部２１に入力され、推定部２２は、この入力した映像を学習モデルＤＢ２３に格納されるモデルに入力する。上記第１のＤＮＮから原映像Ａの特徴である特徴f_1が出力され、第２のＤＮＮから競技者マ原映像Ｂの特徴である特徴f_2が出力される。これらの特徴は推定部２２内のlayersに入力され、このlayersから競技スコアの差の予測値が出力される。

[0037] 以上説明したように、本実施形態に係る映像評価装置では、競技者マスク映像を用いた学習処理により、ヒトの動作に関する映像中の特徴の抽出が促進され、ヒトの動作に汎化する効果が得られるようになり、例えばスポーツ競技における採点を正確に行えるようになる。

また、上記の実施形態では、上記第１、第２および第３のパターンの全てによる学習処理が実行されることでモデルのパラメータが学習される例を示したが、これに限らず、上記第１のパターンの学習処理を必ず含む、第１、第２および第３のパターンの一部による学習処理、すなわち、上記第１のパターンの学習処理が実行され、かつ上記第２および第３のパターンの何れか一方の学習処理が実行されることでモデルのパラメータが学習されるようにしても、上記のヒトの動作に関する映像中の特徴の抽出が促進され、ヒトの動作に汎化する効果が得られる。

上記のように、上記第１のパターンの学習処理を必ず含む理由は、推定処理では、上記のように原映像同士の組み合わせが入力されて予測値が出力されており、学習処理と推定処理とでの入力内容の剥離に起因する、推定の精度の低下を生じさせないためである。

また、上記生ずる効果の観点での、上記のように第１、第２および第３のパターンの全てによる学習処理が実行されたときのモデルのパラメータの学習の精度は、上記のように第１、第２および第３のパターンの一部による学習処理が実行されたときのモデルのパラメータの学習の精度と比較して高いことは勿論である。

[0038] 図７は、本発明の一実施形態に係る画像評価装置のハードウエア構成の一

例を示すブロック図である。

図７に示された例では、上記の実施形態に係る画像評価装置１００は、例えばサーバコンピュータ（server computer）またはパーソナルコンピュータ（personal computer）により構成され、ＣＰＵ（Central Processing Unit）等のハードウエアプロセッサ（hardware processor）１１１Ａを有する。そして、このハードウエアプロセッサ１１１Ａに対し、プログラムメモリ（program memory）１１１Ｂ、データメモリ（data memory）１１２、入出力インタフェース（interface）１１３及び通信インタフェース１１４が、バス（bus）１１５を介して接続される。

[0039] 通信インタフェース１１４は、例えば１つ以上の無線の通信インタフェースユニット（interface unit）を含んでおり、通信ネットワークＮＷとの間で情報の送受信を可能にする。無線インタフェースとしては、例えば無線ＬＡＮ（Local Area Network）などの小電力無線データ通信規格が採用されたインタフェースが使用される。

[0040] 入出力インタフェース１１３には、画像評価装置１００に付設される、利用者などにより用いられる入力デバイス（device）２００および出力デバイス３００が接続される。

入出力インタフェース１１３は、キーボード（keyboard）、タッチパネル（touch panel）、タッチパッド（touchpad）、マウス（mouse）等の入力デバイス２００を通じて利用者などにより入力された操作データを取り込むとともに、出力データを液晶または有機ＥＬ（Electro Luminescence）等が用いられた表示デバイスを含む出力デバイス３００へ出力して表示させる処理を行なうことができる。なお、入力デバイス２００および出力デバイス３００には、画像評価装置１００に内蔵されたデバイスが使用されてもよく、また、ネットワークＮＷを介して画像評価装置１００と通信可能である他の情報端末の入力デバイスおよび出力デバイスが使用されてもよい。

[0041] プログラムメモリ１１１Ｂは、非一時的な有形の記憶媒体として、例えば、ＨＤＤ（Hard Disk Drive）またはＳＳＤ（Solid State Drive）等の随時

書込みおよび読出しが可能な不揮発性メモリ（non-volatile memory）と、ＲＯＭ等の不揮発性メモリとが組み合わせて使用されたもので、一実施形態に係る各種制御処理等を実行する為に必要なプログラムが格納され得る。

[0042] データメモリ１１２は、有形の記憶媒体として、例えば、上記の不揮発性メモリと、ＲＡＭ等の揮発性メモリ（volatile memory）とが組み合わせて使用されたもので、各種処理が行なわれる過程で取得および作成された各種データまたは情報が記憶される為に用いられ得る。

[0043] 本発明の一実施形態に係る画像評価装置１００は、ソフトウエア（software）による処理機能部として、図１に示される各部を有するデータ処理装置として構成され得る。

[0044] 画像評価装置１００の各部によるワークメモリなどとして用いられる情報記憶部は、図７に示されたデータメモリ１１２が用いられることで構成され得る。ただし、これらの構成される記憶領域は画像評価装置１００内に必須の構成ではなく、例えば、ＵＳＢ（Universal Serial Bus）メモリなどの外付け記憶媒体、又はクラウド（cloud）に配置されたデータベースサーバ（database server）等の記憶装置に設けられた領域であってもよい。

[0045] 上記の各部における処理機能部は、いずれも、プログラムメモリ１１１Ｂに格納されたプログラムを上記ハードウエアプロセッサ１１１Ａにより読み出させて実行させることにより実現され得る。なお、これらの処理機能部の一部または全部は、特定用途向け集積回路（ＡＳＩＣ（Application Specific Integrated Circuit））またはＦＰＧＡ（Field-Programmable Gate Array）などの集積回路を含む、他の多様な形式によって実現されてもよい。

[0046] また、各実施形態に記載された手法は、計算機（コンピュータ）に実行させることができるプログラム（ソフトウエア手段）として、例えば磁気ディスク（フロッピー（登録商標）ディスク（Floppy disk）、ハードディスク（hard disk）等）、光ディスク（optical disc）（ＣＤ－ＲＯＭ、ＤＶＤ、ＭＯ等）、半導体メモリ（ＲＯＭ、ＲＡＭ、フラッシュメモリ（Flash memory）等）等の記録媒体に格納し、また通信媒体により伝送して頒布され得る。

なお、媒体側に格納されるプログラムには、計算機に実行させるソフトウエア手段（実行プログラムのみならずテーブル（table）、データ構造も含む）を計算機内に構成させる設定プログラムをも含む。本装置を実現する計算機は、記録媒体に記録されたプログラムを読み込み、また場合により設定プログラムによりソフトウエア手段を構築し、このソフトウエア手段によって動作が制御されることにより上述した処理を実行する。なお、本明細書でいう記録媒体は、頒布用に限らず、計算機内部あるいはネットワークを介して接続される機器に設けられた磁気ディスク、半導体メモリ等の記憶媒体を含むものである。

[0047] なお、本発明は、上記実施形態に限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で種々に変形することが可能である。また、各実施形態は適宜組み合わせて実施してもよく、その場合組み合わせた効果が得られる。更に、上記実施形態には種々の発明が含まれており、開示される複数の構成要件から選択された組み合わせにより種々の発明が抽出され得る。例えば、実施形態に示される全構成要件からいくつかの構成要件が削除されても、課題が解決でき、効果が得られる場合には、この構成要件が削除された構成が発明として抽出され得る。

符号の説明

[0048] １００…画像評価装置

１０…学習処理部

１１，２１…入力部

１２…学習部

１３，２３…学習モデルＤＢ（データベース）

２０…推定処理部

２２…推定部

請求の範囲

[請求項1] 第１の人物の動作に係る動画像および第２の人物の動作に係る動画像を、それぞれの動画像の特徴を求めるモデルに入力し、前記モデルから出力された特徴に基づいて、前記第１の人物の動作に係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作に係る動画像での前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新する第１の更新部と、

前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の少なくとも一方に対して人物に係る領域にマスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新する第２の更新部と、

を備える画像評価装置。

[請求項2] 前記第２の更新部は、

前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の双方に前記マスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値がゼロに近づくように、前記モデルのパラメータを更新する第１の更新処理、および、

前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の一方に前記マスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力され

た特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が、前記それぞれの動画像のうち前記マスク処理が施されない動画像での人物の動作の評価スコアの真値に近づくように、前記モデルのパラメータを更新する第２の更新処理の少なくとも一方を行なう、

請求項１に記載の画像評価装置。

[請求項3] 画像評価装置により行なわれる方法であって、

前記画像評価装置の第１の更新部により、第１の人物の動作に係る動画像および第２の人物の動作に係る動画像を、それぞれの動画像の特徴を求めるモデルに入力し、前記モデルから出力された特徴に基づいて、前記第１の人物の動作に係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作に係る動画像での前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新することと、

前記画像評価装置の第２の更新部により、前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の少なくとも一方に対して人物に係る領域にマスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が真値に近づくように、前記モデルのパラメータを更新することと、

を備える画像評価方法。

[請求項4] 前記第２の更新部は、

前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の双方に前記マスク処理が施されたときのそれぞれの動

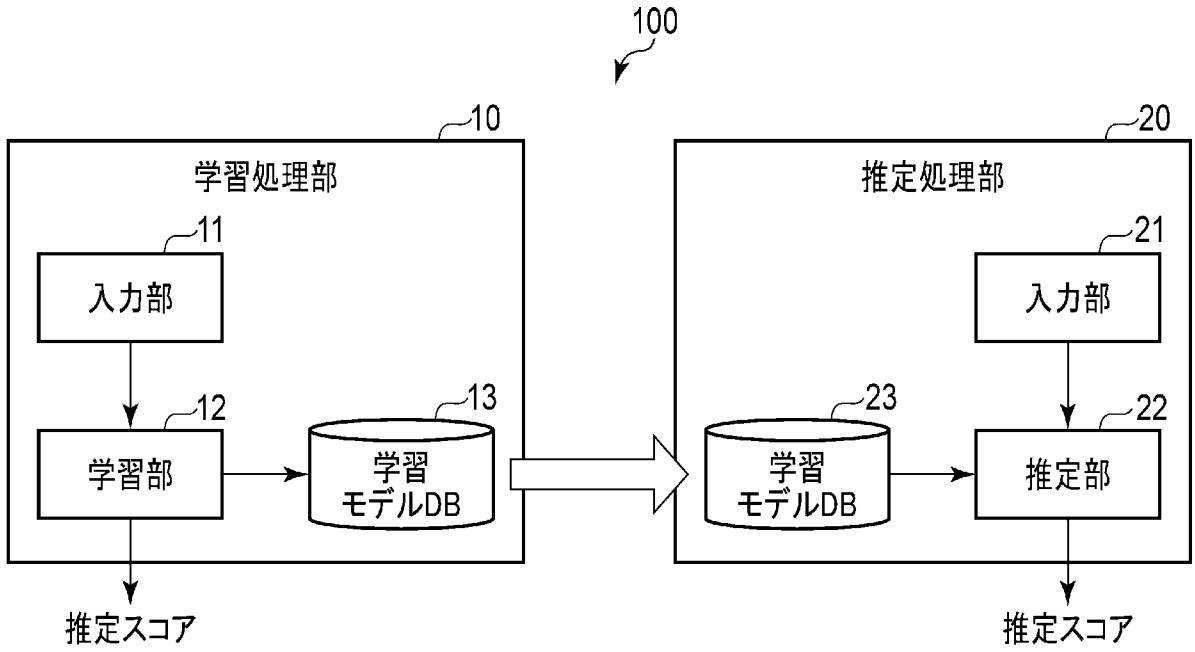
画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値がゼロに近づくように、前記モデルのパラメータを更新する第１の更新処理、および、

前記第１の人物の動作に係る動画像および前記第２の人物の動作に係る動画像の一方に前記マスク処理が施されたときのそれぞれの動画像を前記モデルに入力し、この入力により前記モデルから出力された特徴に基づいて、この出力されたモデルに係る動画像での前記第１の人物の動作の評価スコアおよび前記第２の人物の動作の評価スコアの差分の推定値を計算し、この計算された推定値が、前記それぞれの動画像のうち前記マスク処理が施されない動画像での人物の動作の評価スコアの真値に近づくように、前記モデルのパラメータを更新する第２の更新処理の少なくとも一方を行なう、

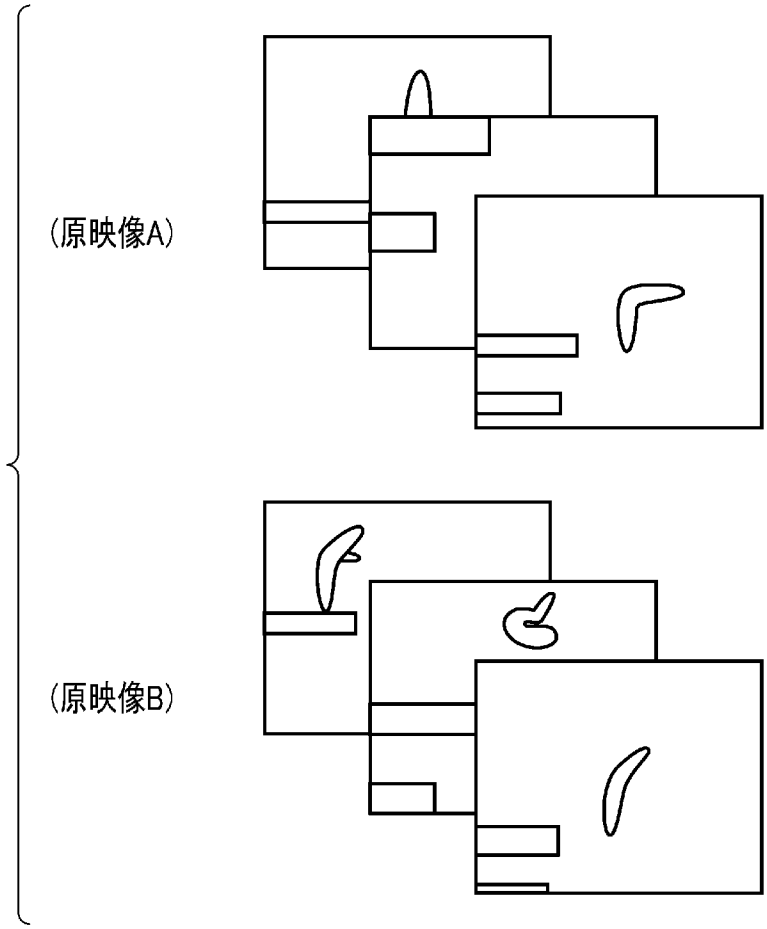
請求項３に記載の画像評価方法。

[請求項5] 請求項１または２に記載の画像評価装置の各部としてプロセッサを機能させる画像評価処理プログラム。

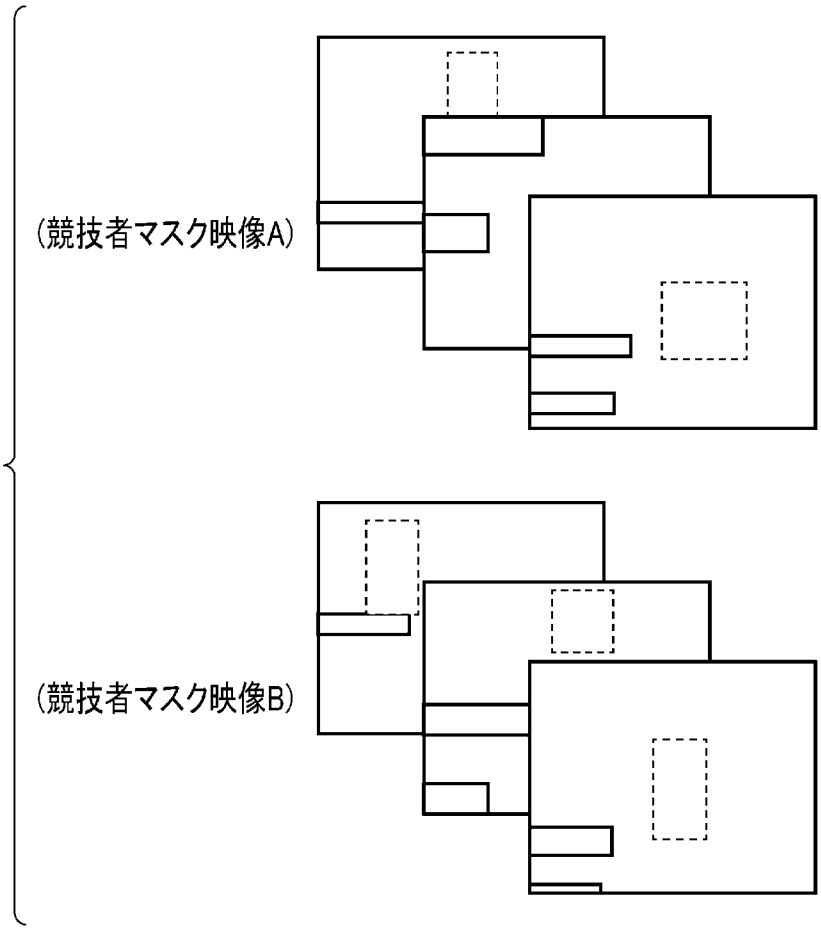
[図1]


100
10
学習処理部
11
入力部
12
学習部
13
学習
モデルDB
推定スコア
20
推定処理部
21
入力部
23
学習
モデルDB
22
推定部
推定スコア

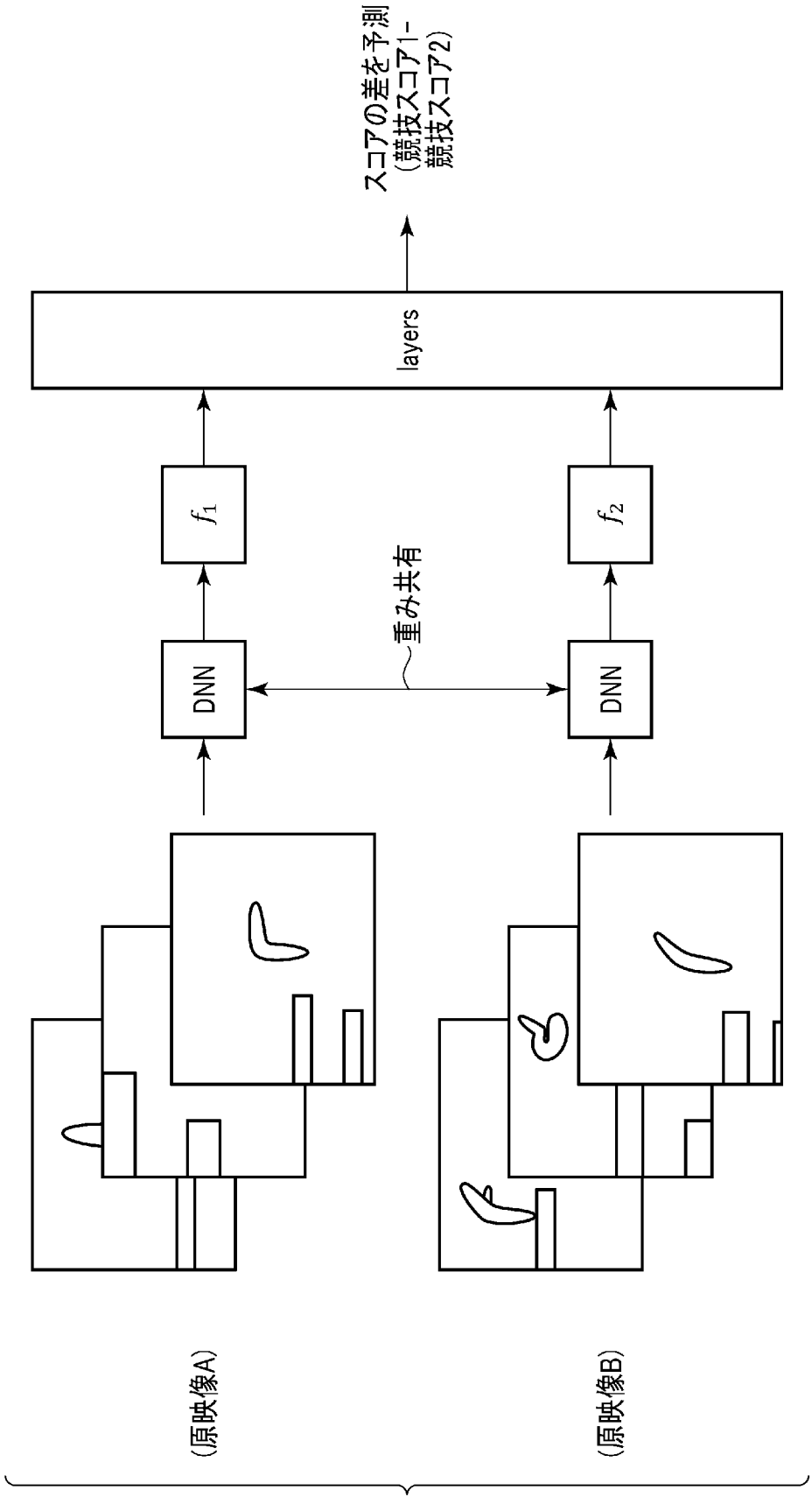
[図2]



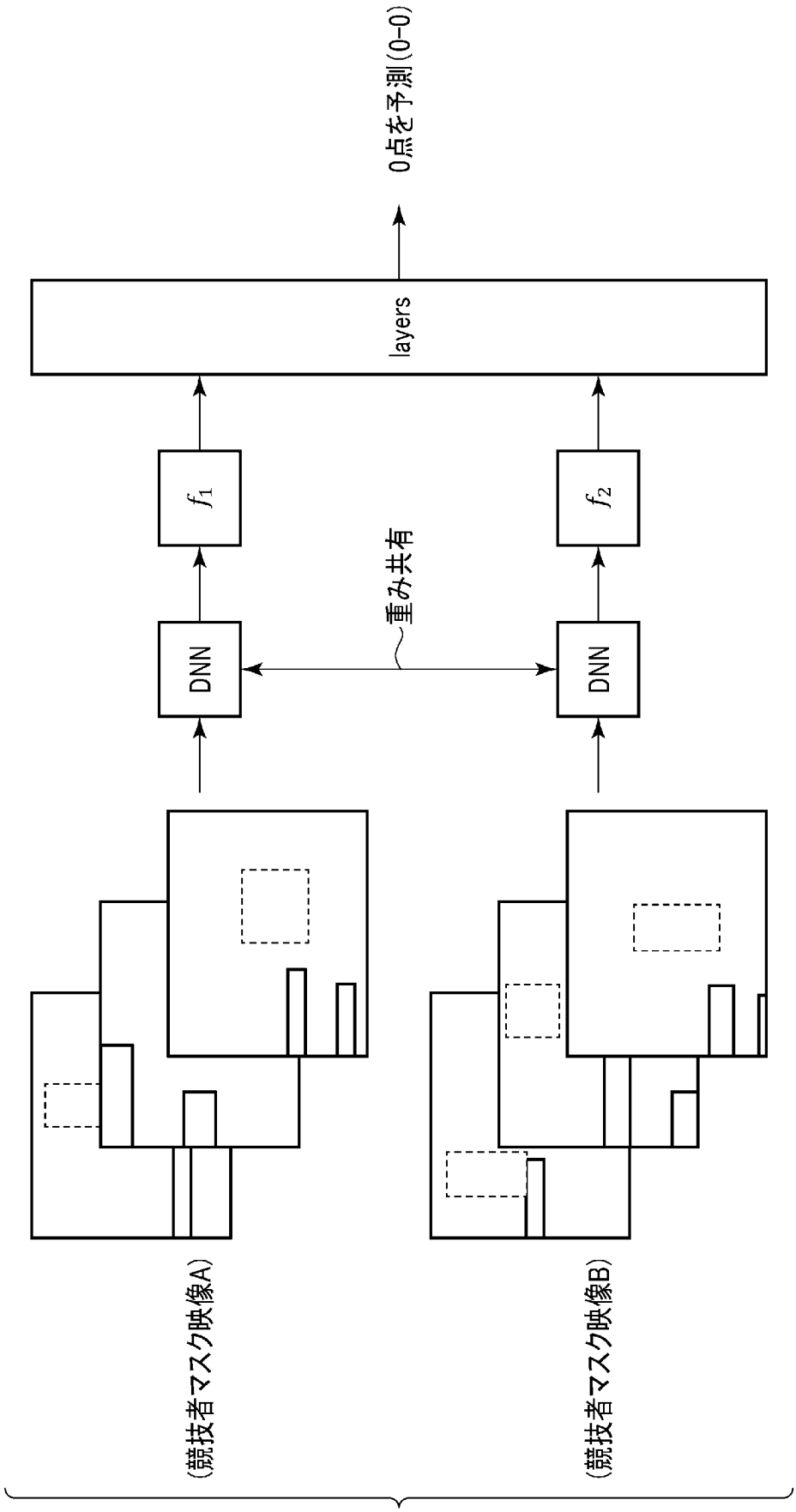
[図3]



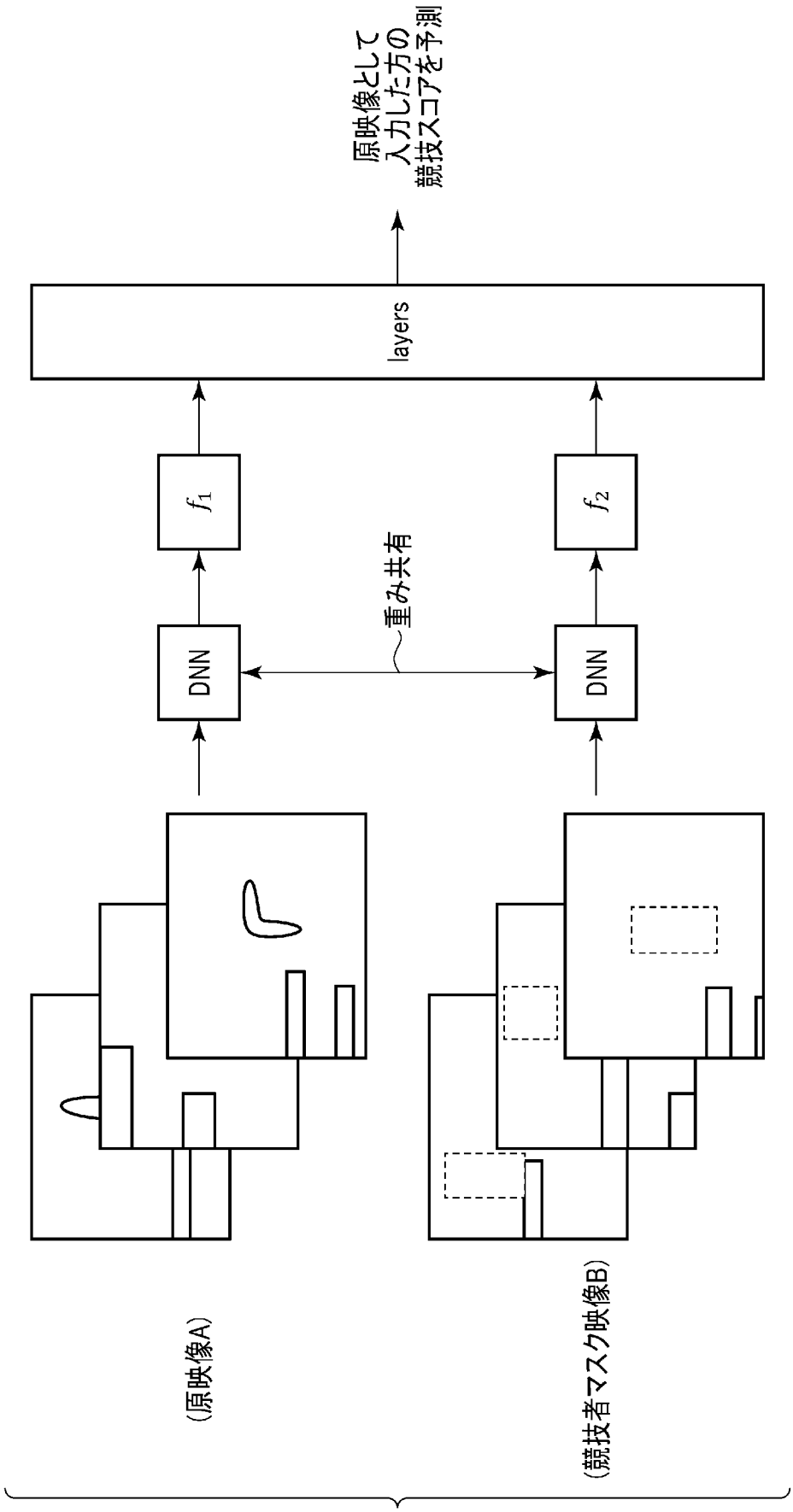
[図4]



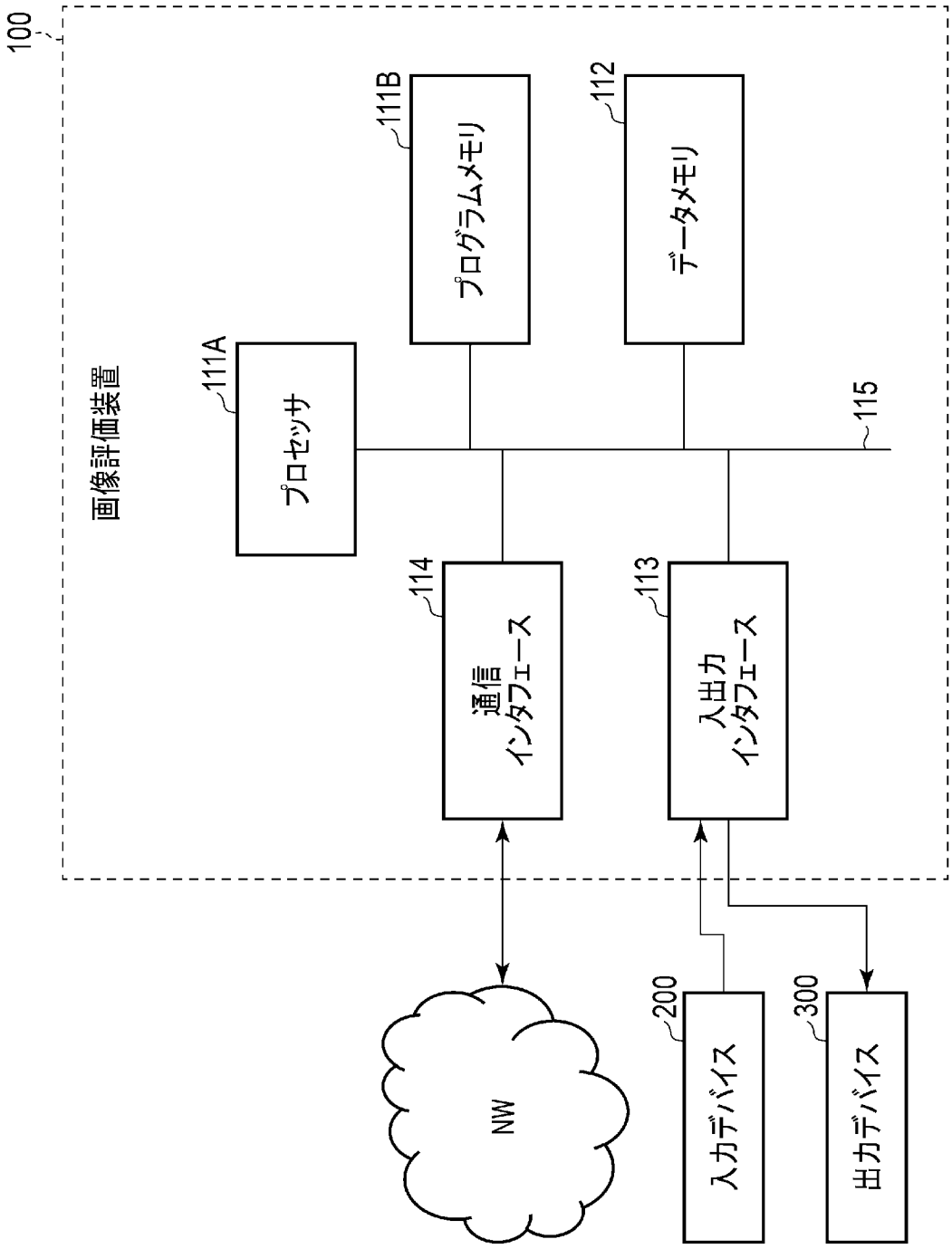
[図5]



[図6]


原映像として
入力した方の
競技スコアを予測
layers
f_1
f_2
重み共有
DNN
DNN
(原映像A)
(競技者マスク映像B)

[図7]


100
画像評価装置
111A
プロセッサ
111B
プログラムメモリ
112
データメモリ
115
114
通信
インタフェース
113
入出力
インタフェース
NW
200
入力デバイス
300
出力デバイス

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/020354

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(2017.01)i; ***G06T 7/20***(2017.01)i
FI: G06T7/20 300Z; G06T7/00 350B

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T7/00; G06T7/20

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2023
Registered utility model specifications of Japan 1996-2023
Published registered utility model applications of Japan 1994-2023

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

JSTPlus/JMEDPlus/JST7580 (JDreamIII); IEEE Xplore; THE ACM DIGITAL LIBRARY

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2022/244135 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 24 November 2022 (2022-11-24) paragraphs [0023]-[0071], fig. 1-5	1-5
Y	CHEN, Ting et al., "A simple framework for contrastive learning of visual representations", Proceedings of the 37th International Conference on Machine Learning, 13 July 2020, no. 149, pp. 1597-1607 [2.1. The Contrastive Learning Framework], fig. 2	1-5
Y	CHEN, Zekai et al., "Masked Image Modeling Advances 3D Medical Image Analysis", 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 02 January 2023, pp. 1969-1979 fig. 1(a)	1-5

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 July 2023	**01 August 2023**
Name and mailing address of the ISA/JP **Japan Patent Office (ISA/JP)** **3-4-3 Kasumigaseki, Chiyoda-ku, Tokyo 100-8915** **Japan**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/020354

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2022/244135 A1	24 November 2022	(Family: none)	

国際調査報告

国際出願番号
PCT/JP2023/020354

A. 発明の属する分野の分類（国際特許分類（ＩＰＣ））

G06T 7/00(2017.01)i; G06T 7/20(2017.01)i
FI: G06T7/20 300Z; G06T7/00 350B

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（ＩＰＣ））

G06T7/00; G06T7/20

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０２３年
日本国実用新案登録公報	１９９６－２０２３年
日本国登録実用新案公報	１９９４－２０２３年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlus/JMEDPlus/JST7580 (JDreamIII); IEEE Xplore; THE ACM DIGITAL LIBRARY

C. 関連すると認められる文献

引用文献のカテゴリー*	引用文献名　及び一部の箇所が関連するときは、その関連する箇所の表示	関連する請求項の番号
Y	WO 2022/244135 A1 (日本電信電話株式会社) 24.11.2022 (2022-11-24) 段落[0023]-[0071], 図1-5	1-5
Y	CHEN, Ting et al., "A simple framework for contrastive learning of visual representations", Proceedings of the 37th International Conference on Machine Learning, 2020.07.13, No.149, pp.1597-1607 [2.1. The Contrastive Learning Framework], 図2	1-5
Y	CHEN, Zekai et al., "Masked Image Modeling Advances 3D Medical Image Analysis", 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023.01.02, pp.1969-1979 図1(a)	1-5

☐ C欄の続きにも文献が列挙されている。

☑ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー
"A" 特に関連のある文献ではなく、一般的技術水準を示すもの
"E" 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
"L" 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
"O" 口頭による開示、使用、展示等に言及する文献
"P" 国際出願日前で、かつ優先権の主張の基礎となる出願の日の後に公表された文献
"T" 国際出願日又は優先日後に公表された文献であって出願と抵触するものではなく、発明の原理又は理論の理解のために引用するもの
"X" 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
"Y" 特に関連のある文献であって、当該文献と他の１以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
"&" 同一パテントファミリー文献

国際調査を完了した日	国際調査報告の発送日
18.07.2023	01.08.2023

名称及びあて先
日本国特許庁(ISA/JP)
〒100-8915
日本国
東京都千代田区霞が関三丁目４番３号

権限のある職員（特許庁審査官）
佐田　宏史 5H 4189
電話番号 03-3581-1101 内線 3531

国際調査報告
パテントファミリーに関する情報

国際出願番号
PCT/JP2023/020354

引用文献	公表日	パテントファミリー文献	公表日
WO 2022/244135 A1	24.11.2022	（ファミリーなし）	