

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7553266号
(P7553266)

(45)発行日 令和6年9月18日(2024.9.18)

(24)登録日 令和6年9月9日(2024.9.9)

(51)国際特許分類

F I

G 0 6 T 7/00 (2017.01) G 0 6 T 7/00 3 5 0 C

G 0 6 N 3/02 (2006.01) G 0 6 N 3/02

請求項の数 12 外国語出願 (全59頁)

(21)出願番号	特願2020-81176(P2020-81176)	(73)特許権者	500102435
(22)出願日	令和2年5月1日(2020.5.1)		ダッソー システムズ
(65)公開番号	特開2020-205033(P2020-205033 A)		D A S S A U L T S Y S T E M E S
(43)公開日	令和2年12月24日(2020.12.24)		フランス国 7 8 1 4 0 ペリジー ピラ
審査請求日	令和5年4月3日(2023.4.3)		クブレー リュ マルセル ダッソー 1 0
(31)優先権主張番号	19305580.3	(74)代理人	110000752
(32)優先日	令和1年5月6日(2019.5.6)		弁理士法人朝日特許事務所
(33)優先権主張国・地域又は機関	欧州特許庁(EP)	(72)発明者	レジェブ・スファー アズマ
			フランス国 7 8 1 4 0 ヴェリジーヴィ
			ラクブレー マルセルダッソー通り1 0
			ダッソーシステムズ社内
		(72)発明者	ボルケナフド マリカ
			フランス国 7 8 1 4 0 ヴェリジーヴィ
			ラクブレー マルセルダッソー通り1 0
			ダッソーシステムズ社内
			最終頁に続く

(54)【発明の名称】 仮想世界での体験学習

(57)【特許請求の範囲】

【請求項1】

機械学習のコンピュータ実施方法であって、
テストドメインに属するシーンのテストデータセット（S 1 0）、及び
ドメイン適応型ニューラルネットワークは訓練ドメインから取得されたデータ上で学習された機械学習型ニューラルネットワークであり、当該テストドメインのシーンにおける空間的に再構成可能なオブジェクトの推論のために構成されるドメイン適応型ニューラルネットワーク（S 4 0）
を提供するステップと、
データ分布に関してテストドメインよりも訓練ドメインに近い中間ドメインを決定するステップ（S 5 0）と、
ドメイン適応ニューラルネットワークを適用することによって、中間ドメイン上で転送されたテストドメインのシーンから空間的に再構成可能なオブジェクトを推論するステップ（S 6 0）と
を有し、
前記中間ドメインを決定するステップ（S 5 0）は、前記テストデータセットの各シーンについて、
テストデータセットのシーンを、データ分布に関して訓練ドメインにより近い別のシーンに変換するステップ
を含み、

10

前記シーンの変換は、テストデータセットのシーンから仮想シーンを生成するステップ（S 5 0 0）を含み、

別のシーンは、テストデータセットのシーンおよび生成された仮想シーンに基づいて推論され、

前記中間ドメインを決定するステップ（S 5 0）は、前記テストデータセットのシーンと前記生成された仮想シーンを混合するステップ（S 5 1 0）をさらに含み、前記混合するステップは前記別のシーンをもたらす

コンピュータ実施方法。

【請求項 2】

前記テストデータセットのシーンから前記仮想シーンを生成するステップ（S 5 0 0）は、前記テストデータセットのシーンに仮想シーンジェネレータを適用するステップを含み、

10

前記仮想シーンジェネレータは、前記テストドメインの前記シーンから前記仮想シーンを推論するように構成される

機械学習ニューラルネットワークで請求項 1 に記載の方法。

【請求項 3】

前記仮想シーンジェネレータは、各々が 1 つ以上の空間的に再構成可能なオブジェクトを含むシーンのデータセットについて学習済である

請求項 2 に記載の方法。

【請求項 4】

20

前記テストデータセットのシーンと前記生成された仮想シーンを混合するステップ（S 5 1 0）は、線形混合である

請求項 1 に記載の方法。

【請求項 5】

前記テストデータセットは、実シーンを含む

請求項 1 乃至 4 のいずれか一項に記載の方法。

【請求項 6】

前記テストデータセットの各現実シーンは 1 つ以上の空間的に再構成可能な製造ツールを含む実製造シーンであり、前記ドメイン適応型ニューラルネットワークは、実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される

30

請求項 5 に記載の方法。

【請求項 7】

前記訓練ドメインは、各々が 1 つ以上の空間的に再構成可能なオブジェクトを含む仮想シーンの訓練データセットを含む

請求項 1 乃至 6 のいずれか一項に記載の方法。

【請求項 8】

前記仮想シーンのデータセットの各仮想シーンは 1 つ以上の空間的に再構成可能な製造ツールを含む仮想製造シーンであり、前記ドメイン適応型ニューラルネットワークは、実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される

請求項 7 に記載の方法。

40

【請求項 9】

前記訓練ドメインから取得された前記データは別の中間ドメインのシーンを含み、前記ドメイン適応ニューラルネットワークは別の中間ドメイン上で学習済であり、前記別の中間ドメインはデータ分布に関して前記訓練ドメインよりも前記中間ドメインに近い

請求項 1 乃至 8 のいずれか一項に記載の方法。

【請求項 10】

請求項 1 乃至 9 のいずれか一項に記載の方法を実行するための命令を含むコンピュータプログラム。

【請求項 11】

請求項 10 に記載のコンピュータプログラムを記録したデータ記憶媒体を含む装置。

50

【請求項 1 2】

前記データ記憶媒体に結合されたプロセッサをさらに備える請求項 1 1 に記載の装置。

【発明の詳細な説明】

【技術分野】

【0 0 0 1】

本開示はコンピュータプログラム及びシステムの分野に関し、より具体的には、空間的に再構成可能なオブジェクトを含むシーンのデータセットを用いた機械学習の方法、システム、及びプログラムに関する。

【背景技術】

【0 0 0 2】

物体の設計、エンジニアリング及び製造のために、多数のシステム及びプログラムが市場に提供されている。C A D はコンピュータ支援設計の頭字語であり、例えば、オブジェクトを設計するためのソフトウェアソリューションに関する。C A E はコンピュータ支援エンジニアリングの頭字語であり、例えば、将来の製品の物理的挙動をシミュレートするためのソフトウェアソリューションに関連する。C A M はコンピュータ支援製造の頭字語であり、例えば、製造処理及び動作を定義するためのソフトウェアソリューションに関する。そのようなコンピュータ支援設計システムでは、グラフィカルユーザーインターフェースが技法の効率性に関して重要な役割を果たす。これらの技術は、製品ライフサイクル管理 (P L M) システム内に組み込まれてもよい。P L M とは、企業が製品データを共有し、共通の処理を適用し、企業知識を活用して、長期的な企業のコンセプトを越えて、コンセプトから生涯にわたる製品の開発に役立てることを支援するビジネス戦略のことをいう。(C A T I A 、 E N O V I A 及び D E L M I A の商標で) ダッソーシステムズによって提供される P L M ソリューションは、製品工学知識を編成するエンジニアリングハブと、製造工学知識を管理する製造ハブと、企業がエンジニアリングハブ及び製造ハブの両方に接続することを可能にする企業ハブとを提供する。全体として、システムは最適化された製品定義、製造準備、生産、及びサービスを駆動する動的な知識ベースの製品作成及び意思決定サポートを可能にするために、製品、プロセス、リソースをリンクするオープンオブジェクトモデルを提供する。

【0 0 0 3】

このコンテキスト及び他のコンテキストにおいて、機械学習は、広く重要性を増している。

【0 0 0 4】

以下の論文はこの分野に関するものであり、本稿において参照される。

- [1] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in CVPR, 2016.
- [2] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in ECCV, 2016.
- [3] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in ICRA, 2017.
- [4] S. R. Richter, Z. Hayder, and V. Koltun, "Playing for benchmarks," in ICCV, 2017.
- [5] S. Hinterstoisser, V. Lepetit, P. Wohlhart, and K. Konolige, "On pretrained image features and synthetic images for deep learning," in arXiv:1710.10710, 2017.
- [6] D. Dwibedi, I. Misra, and M. Hebert, "Cut, paste and learn: Surprisingly easy synthesis for instance detection," in ICCV, 2017.
- [7] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in IEEE/RSJ International Conference on Intelligent Robots and S

10

20

30

40

50

systems (IROS), 2017.

[8] J. Tremblay, A. Prakash, D. Acuna, M. Brophy, V. Jampani, C. Anil, T. To, E. Cameracci, S. Bochoon, and S. Birchfield, "Training deep networks with synthetic data: Bridging the reality gap by domain randomization," in CVPR Workshop on Autonomous Driving (WAD), 2018.

[9] K. He, G. Gkioxari, P. Dollar, and R. Girshick. "Mask r-cnn". arXiv:1703.06870, 2017.

[10] J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ILSVRC-2017, 2017. URL <http://www.image-net.org/challenges/LSVRC/2017/>.

[11] Y. Chen, W. Li, and L. Van Gool. ROAD: Reality oriented adaptation for semantic segmentation of urban scenes. In CVPR, 2018.

[12] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool. Domain adaptive Faster R-CNN for object detection in the wild. In CVPR, 2018.

[13] Y. Zou, Z. Yu, B. Vijaya Kumar, and J. Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In ECCV, September 2018.

[14] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in CVPR, 2012.

[15] A. Prakash, S. Bochoon, M. Brophy, D. Acuna, E. Cameracci, G. State, O. Shapira, and S. Birchfield. Structured domain randomization: Bridging the reality gap by context-aware synthetic data. arXiv preprint arXiv:1810.10093, 2018.

[16] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

[17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollar, and C. L. Zitnick. Microsoft COCO: Common objects in context. In ECCV. 2014.

[18] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. CoRR, abs/1506.01497, 2015.

[19] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation, Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations, MIT Press, Cambridge, MA, 1986.

[20] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. CoRR, abs/1703.10593, 2017.

[21] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Advances in Neural Information Processing Systems (NIPS), 2017.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1-9, 2015.

【 0 0 0 5 】

深層学習技術は、2D / 3D物体認識及び検出、意味的セグメント化、ポーズ推定、又はモーションキャプチャ（[9, 10]を参照）などのいくつかの分野で顕著な性能を示している。しかしながら、そのような技術は通常、（1）それらの完全な可能性で実行するために膨大な量の訓練データを必要とし、（2）しばしば、高価な手動ラベル付け（アノテーションとしても知られる）を必要とする。

【 0 0 0 6 】

（1）何十万ものラベル付けされたデータを収集することは実際に、製造、健康、又は

10

20

30

40

50

屋内のモーションキャプチャの分野などのいくつかの分野ではデータの不足又は機密性の理由のために、非常に困難である可能性がある。公的に利用可能なデータセットは、通常、クラウドソーシングプラットフォームを介して手動でラベル付けされた日常生活のオブジェクトを含む（[1 6 , 1 7] 参照）。

【 0 0 0 7 】

（ 2 ）データ収集問題に加えて、ラベリングは非常に時間がかかる。よく知られている ImageNet データセット（[1 6] 参照）のアノテーション（各画像は「唯一の」単一ラベルによってラベル付けされる）は、何年も続いている。ピクセル単位のラベリングでは、画像あたり平均 1 時間以上必要である。また、手動アノテーションは、いくつかのエラー及び正確さの欠如を提示する可能性がある。いくつかのアノテーションはドメイン固有のノウハウ（例えば、製造装置のラベル付け）を必要とする。

10

【 0 0 0 8 】

これら全ての理由から、近年、深層モデルを学習するために仮想データを利用することが注目されている（[1 , 2 , 3 , 4 , 5 , 6 , 7 , 8 , 1 5] 参照）。仮想（合成とも呼ばれる）データは、実際のイベントによって生成されるのではなく、CAD 又は 3 D モデリングソフトウェアを使用して作成されるコンピュータ生成データである。実際、仮想データは、既存の（現実）データでは利用できない非常に具体的なニーズや条件を満たすために生成されることができる。これは、プライバシーがデータの可用性や使用を制限する必要がある場合や、テスト環境に必要なデータが単に存在しない場合に役立つ。仮想データでは、ラベル付けは費用がかからず、エラーがないことに留意されたい。

20

【 0 0 0 9 】

しかしながら、仮想データと現実データとの間の固有のドメイン差（現実ギャップとしても知られる）は、学習モデルが現実世界のシナリオに適格であることを妨げる可能性がある。これは、主に、オーバーフィッティングの課題によるものであり、ネットワークは仮想データにのみ存在する詳細を学習し、十分に一般化することができず、実データの有益な表現を抽出することができない。

【 0 0 1 0 】

最近普及している 1 つの方法は、特に自律車に関する認識問題に対して、ビデオゲームデータのような、低コストで自動的にアノテーションを付けることができるフォトリアリスティックな仮想データを収集することである。

30

【 0 0 1 1 】

例えば、Richter ら（[2] 参照）は、GTAV ゲームからの意味的セグメンテーションのための大規模合成都市シーンデータセットを構築した。GTAV ベースのデータ（[2 , 3 , 4] 参照）は、多数のアセットを使用して、ピクセルレベル（~ 2 5 K 画像）で自動的にラベル付けすることができる様々な現実的な運転環境を生成する。そのようなフォトリアリスティックな仮想データはしばしば、運転及び都市の状況以外のドメインには存在しない。例えば、既存の仮想製造環境は、非フォトリアリスティック CAD ベースの環境である（製造ドメインに対してビデオゲームは存在しない）。製造仕様ではフォトリアリズムは不要である。さらに、このような仮想環境では、シェーディング及びライティングを有すると混乱する可能性がある。したがって、すべての努力は、フォトリアリスティックな CAD モデルではなく、意味論的及び幾何学的に正確な CAD モデルを作ることに集中している。さらに、異なる工場と工場のレイアウトとの間の高い変動性のために、訓練目的のためのフォトリアリスティックな製造環境を作ることは適切ではない。都市シーンが強い特異性（例えば、建物、街路、車等のサイズ及び空間的關係）を有する都市コンテキストとは異なり、シーンを製造するための「固有の」空間構造は存在しない。また、VKITTI（[1] 参照）のようないくつかの以前の研究は現実世界の運転シーンのレプリカを生成し、したがって、それらの元の現実データと高度に相関し、変動性を欠いていたことにも留意されたい。

40

【 0 0 1 2 】

さらに、高レベルの現実感があっても、現実のデータに作用するようにニューラルネッ

50

トワークを訓練するために、フォトリアリスティックなデータをどのように有効に使用するかは明らかではなかった。多くの場合、仮想表現と実表現との間のギャップを低減するために、ニューラルネットワークを訓練する間に、クロスドメイン適応成分（[1 1 , 1 2 , 1 3]を参照）が使用される。

【 0 0 1 3 】

要約すると、フォトリアリスティック仮想データセットは、注意深く設計されたシミュレーション環境、又は製造などの多くの分野では実現不可能な出発点としてのアノテーション付き実データの存在のいずれかを必要とする。

【 0 0 1 4 】

このような困難性を軽減するために、T o b i nら（[7]参照）は、ランダムな変化を優先して現実的なレンダリングが回避されるドメインランダム化（D R）の概念を導入した。彼らのアプローチは、前景オブジェクトのテクスチャと色、背景イメージ、シーン内の光の数、光のポーズ、カメラ位置、前景オブジェクトをランダムに変化させる。目標は、ネットワークが実世界データを単に別のパリエーションとして見るのに十分なパリエーションを有する仮想データを生成することによって、現実のギャップを閉じることである。D Rを使用して、彼らは、テーブルに固定されたロボットアームに対する様々な形状ベースの物体の3 D世界位置を推定するためにニューラルネットワークを訓練した。最近の研究（[8]参照）は、現実世界のK I T T Iデータセットにおける自動車の2 Dバウンディングボックス検出において最新技術を達成するD Rの能力を実証した（[1 4]参照）。しかしながら、D Rは変動量が与えられると、訓練するのに非常に大量のデータ（オブジェクトクラス当たり~ 1 0 0 K）を必要とし、多くの場合、ネットワークは正しい特徴を学習することが困難であることを見出し、コンテキストの欠如は、D Rが小さい又は遮蔽されたオブジェクトを検出することを妨げる。実際、その作業の結果は、（1）完全に見える（K I T T I E a s y d a t a s e t）大型車に限られていた。オブジェクト検出などの複雑な推論タスクに使用される場合、このアプローチは学習モデルが周囲のコンテキストなしで決定を行うために、境界ボックス内に十分なピクセルを必要とする。これらの欠点は全て、定量的に確認された（[1 5]参照）。

【 0 0 1 5 】

より困難な基準（例えば、より小さな車、部分的閉塞）を扱うために、[1 5]の著者らはD Rの変種を提案した。そこでは以前のD Rベースの研究で使用された一様な確率分布の代わりに、手元の特定の問題（すなわち、都市コンテキストにおける自動車検出）から生じる確率分布に従って、オブジェクトをランダムに配置することにより、都市のコンテキスト情報と都市構造シーンを利用する。

【 0 0 1 6 】

また、D Rベースの方法（[5、6、7、1 5]参照）は主に、クラス内変動性がほとんどない「容易である」オブジェクトクラス（例えば、「自動車」クラス）に焦点を当てている。重要なクラス内変動性は、データ作成中に考慮すべき変動の量が指数関数的に増加する可能性があるため、これらの手法の著しい性能低下を引き起こす可能性がある。これは、空間的に再構成可能なオブジェクト、例えば関節オブジェクトの場合である。

【 0 0 1 7 】

空間的に再構成可能なオブジェクト又は接続オブジェクトは関節を介して取り付けられた成分を有し、各オブジェクトタイプに無限範囲の可能なステータスを与えることができる互いに対して移動することができるオブジェクトである。

【 0 0 1 8 】

同じ問題が、複数のオブジェクトクラス（クラス間及びクラス内変動性）を処理する間に発生する可能性がある。

【 0 0 1 9 】

このコンテキスト内で、空間的に再構成可能なオブジェクトを含むシーンのデータセットを用いて機械学習する改善された方法が依然として必要とされている。

【発明の概要】

10

20

30

40

50

【0020】

したがって、本発明は、機械学習のコンピュータ実施方法を提供する。この方法は、シーンのテストデータセットを提供することを含む。テストデータセットは、テストドメインに属する。この方法は、ドメイン適応ニューラルネットワークを提供することを含む。ドメイン適応ニューラルネットワークは、訓練ドメインから得られたデータについて学習された機械学習ニューラルネットワークである。ドメイン適応ニューラルネットワークは、テストドメインのシーンにおける空間的に再構成可能なオブジェクトの推論のために構成される。この方法は、中間ドメインを決定することをさらに含む。中間ドメインはデータ分布の観点から、テストドメインよりも訓練ドメインに近い。この方法は、ドメイン適応ニューラルネットワークを適用することによって、中間ドメイン上で転送されたテストドメインのシーンから空間的に再構成可能なオブジェクトを推論することをさらに含む。

10

【0021】

この方法は、以下のうちの1つ又は複数を含んでもよい。

- ・中間ドメインの決定は、テストデータセットの各シーンについて以下を含む。
 - ・テストデータセットのシーンを、データ分布の観点から訓練ドメインにより近い別のシーンに変換するステップ。
 - ・シーンの変換は以下を含む。
 - ・テストデータセットのシーンから仮想シーンを生成する。別のシーンは、テストデータセットのシーンと生成された仮想シーンに基づいて推論される。
 - ・テストデータセットのシーンから仮想シーンを生成するステップは仮想シーンジェネレータをテストデータセットのシーンに適用することを含み、仮想シーンジェネレータはテストドメインのシーンから仮想シーンを推論するように構成された機械学習ニューラルネットワークである。
 - ・仮想シーンジェネレータは、各々が1つ以上の空間的に再構成可能なオブジェクトを含むシーンのデータセット上で学習済である。
 - ・中間ドメインを決定するステップはテストデータセットのシーンと生成された仮想シーンを混合するステップをさらに含み、混合するステップは別のシーンをもたらす。
 - ・テストデータセットのシーンと生成された仮想シーンとの混合は、線形混合である。
 - ・テストデータセットは実シーンで構成される。
 - ・テストデータセットの各実シーンは1つ以上の空間的に再構成可能な製造ツールを含む実製造シーンであり、ドメイン適応型ニューラルネットワークは、実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。
 - ・訓練ドメインは、各々が1つ以上の空間的に再構成可能なオブジェクトを含む仮想シーンの訓練データセットを含む。
 - ・仮想シーンのデータセットの各仮想シーンは1つまたは複数の空間的に再構成可能な製造ツールを備える仮想製造シーンであり、ドメイン適応型ニューラルネットワークは実製造シーンにおいて空間的に再構成可能な製造ツールを推論するように構成され、及び/又は訓練ドメインから取得されたデータは別の中間ドメインのシーンを含み、ドメイン適応型ニューラルネットワークは別の中間ドメイン上で学習されており、別の中間ドメインはデータ分布に関して訓練ドメインよりも中間ドメインに近い。

20

30

40

【0022】

本方法を実行するための命令を含むコンピュータプログラムがさらに提供される。

【0023】

さらに、コンピュータプログラムが記録されたデータ記憶媒体を含む装置が提供される。

【0024】

この装置は、例えばSaaS、又は他のサーバ、あるいはクラウドベースのプラットフォームなど上で、一時的でないコンピュータ可読媒体を形成するか、又はそのような媒体として機能することができる。デバイスは代替的に、データ記憶媒体に結合されたプロセッサを備えることができる。したがって、装置は全体的に又は部分的にコンピュータシステムを形成することができる(たとえば、装置は、システム全体のサブシステム)。システムは

50

、プロセッサに結合されたグラフィカルユーザーインターフェースをさらに備えることができる。

【図面の簡単な説明】

【 0 0 2 5 】

【図 1】方法の一例を統合したコンピュータ実施プロセスのフローチャートを示す。

【図 2】コンピュータ実装プロセスの例及び / 又は方法の例を示すフローチャート。

【図 3】コンピュータ実装プロセスの例及び / 又は方法の例を示すフローチャート。

【図 4】コンピュータ実装プロセスの例及び / 又は方法の例を示すフローチャート。

【図 5】システムのグラフィカルユーザーインターフェースの一例を示す。

【図 6】システムの一部を示す。

10

【図 7】プロセス及び / 又は方法を示す。

【図 8】プロセス及び / 又は方法を示す。

【図 9】プロセス及び / 又は方法を示す。

【図 10】プロセス及び / 又は方法を示す。

【図 11】プロセス及び / 又は方法を示す。

【図 12】プロセス及び / 又は方法を示す。

【図 13】プロセス及び / 又は方法を示す。

【図 14】プロセス及び / 又は方法を示す。

【図 15】プロセス及び / 又は方法を示す。

【図 16】プロセス及び / 又は方法を示す。

20

【図 17】プロセス及び / 又は方法を示す。

【図 18】プロセス及び / 又は方法を示す。

【図 19】プロセス及び / 又は方法を示す。

【図 20】プロセス及び / 又は方法を示す。

【図 21】プロセス及び / 又は方法を示す。

【図 22】プロセス及び / 又は方法を示す。

【図 23】プロセス及び / 又は方法を示す。

【図 24】プロセス及び / 又は方法を示す。

【図 25】プロセス及び / 又は方法を示す。

【図 26】プロセス及び / 又は方法を示す。

30

【図 27】プロセス及び / 又は方法を示す。

【図 28】プロセス及び / 又は方法を示す。

【図 29】プロセス及び / 又は方法を示す。

【図 30】プロセス及び / 又は方法を示す。

【図 31】プロセス及び / 又は方法を示す。

【図 32】プロセス及び / 又は方法を示す。

【発明を実施するための形態】

【 0 0 2 6 】

本発明は、空間的に再構成可能なオブジェクトを含むシーンのデータセットを用いた機械学習のコンピュータ実施方法を提供する。

40

【 0 0 2 7 】

本発明は特に、機械学習の第 1 の方法を提供する。第 1 の方法は、仮想シーンのデータセットを提供することを含む。仮想シーンのデータセットは、第 1 のドメインに属する。第 1 の方法は、実シーンのテストデータセットを提供することをさらに含む。テストデータセットは、第 2 のドメインに属する。第 1 の方法は、第 3 のドメインを決定することをさらに含む。第 3 のドメインはデータ分布の観点から、第 1 のドメインよりも第 2 のドメインに近い。第 1 の方法は、第 3 のドメインに基づいてドメイン適応ニューラルネットワークを学習することをさらに含む。ドメイン適応ニューラルネットワークは、現実のシーンにおける空間的に再構成可能なオブジェクトの推論のために構成されたニューラルネットワークである。以下、この第 1 の方法を「ドメイン適応学習法」と呼ぶことがある。

50

【 0 0 2 8 】

ドメイン適応学習方法は、空間的に再構成可能なオブジェクトを含むシーンのデータセットを用いて機械学習の改善された方法を形成する。

【 0 0 2 9 】

特に、ドメイン適応学習方法は、現実の場面において空間的に再構成可能なオブジェクトを推論することができるドメイン適応ニューラルネットワークを提供する。これにより、実製造シーンにおける空間的に再構成可能な物体の推論プロセス、例えば、空間的に再構成可能な製造ツール（例えば、産業用多関節ロボット）の推論プロセスにおいて、ドメイン適応学習法を使用することができる。

【 0 0 3 0 】

さらに、ドメイン適応型ニューラルネットワークは現実のシーンのデータセット上で直接学習することはできないが、ドメイン適応型ニューラルネットワークは現実のシーンにおいて現実の空間的に再構成可能なオブジェクトを推論することができる。このような改善は特に、ドメイン適応学習方法に従って第3のドメインを決定することによって達成される。実際に、第1のドメインは仮想シーンで作られるので、データ分布に関して第2のドメイン（すなわち、実シーンで作られるドメイン適応型ニューラルネットワークのテストデータセットのドメイン）から比較的遠くにあってもよい。このような現実のギャップは、仮想シーンのデータセットに直接基づくドメイン適応型ニューラルネットワークの学習を困難及び／又は不正確にする可能性がある。このような学習は実際に、現実のシーンにおける空間的に再構成可能なオブジェクトの推論に関して不正確な結果を生成し得るドメイン適応型ニューラルネットワークをもたらす可能性がある。代わりに、ドメイン適応学習方法はデータ分布に関して、第1のドメインよりも第2のドメインに近い第3のドメインを決定し、この第3のドメイン上でドメイン適応ニューラルネットワークの学習に基づく。これは、提供される（すなわち、ドメイン適応学習方法への入力として与えられる）唯一のデータセットが実際に仮想シーンのデータセットである間に、実シーンにおいて空間的に再構成可能なオブジェクトを正確に推論することができるドメイン適応ニューラルネットワークの学習を可能にする。言い換えれば、学習及び学習されたドメイン適応ニューラルネットワークの両方が、よりロバストである。

【 0 0 3 1 】

これは仮想シーンのデータセットが理論的にはドメイン適応ニューラルネットワークの訓練セット（又は学習セット）であることを意図していること、及び、ドメイン適応学習方法が実際の学習の前に、前述の改善を用いてドメイン適応ニューラルネットワークを学習するように、訓練セットの前処理の段階、すなわち第3のドメインの決定を含むことを意味する。言い換えれば、ドメイン適応ニューラルネットワークの真の訓練／学習ドメインは第3のドメインであり、元々ドメイン適応ニューラルネットワークの訓練／学習ドメイン（すなわち、第1のドメイン）であることが意図されたドメインの前処理に由来する。それにもかかわらず、理論的訓練セット（すなわち、前述のように、訓練に使用される前に処理される）として仮想シーンのデータセットに依存することにより、多数のシーンを有するデータセットを使用することが可能になる。実際、一般に、空間的に再構成可能な製造ツールを含む製造シーンのような、空間的に再構成可能なオブジェクトを含む実シーンの利用可能な大きなデータセットは存在しないか、又は少なくともほとんど存在しない。そのようなデータセットはプライバシー／機密性の問題、例えば、ロボットのような空間的に再構成可能な製造ツールを有する製造シーンは典型的には機密であり、したがって、公に利用可能ではないため、実際に入手することが困難である。さらに、たとえ現実のシーンのそのような大きなデータセットが入手可能であっても、多くのアノテーション（例えば、ラベル付け）の誤りを含む可能性があり、空間的に再構成可能な物体（例えば、ロボットなどの製造ツールなど）にアノテーションを付けることは、空間的に再構成可能な物体が現実のシーンにおけるその認識及び手動によるそのアノテーションが特定の量の技能及び知識を必要とするほど多数及び／又は非常に複雑な位置を有する可能性があるため、そのために適切な技能を有していない人にとって実際に困難である可能性がある。

10

20

30

40

50

一方、仮想シーンの大きなデータセットは、例えばシミュレーションソフトウェアを使用して仮想シーンを容易に生成することによって容易に得ることができる。さらに、空間的に再構成可能なオブジェクトのような仮想シーンにおけるオブジェクトのアノテーション付けは、比較的容易であり、とりわけ、自動的に、間違いなく実行することができる。実際、シミュレーションソフトウェアは典型的にはシミュレーションに関するオブジェクトの仕様を知っており、したがって、それらのアノテーション付けは、ソフトウェアによって自動的に、アノテーション付けミスの危険因子を低くして実行することができる。

【0032】

例では、仮想シーンのデータセットの各仮想シーンが1つ又は複数の空間的に再構成可能な製造ツールを備える仮想製造シーンである。これらの例では、ドメイン適応型ニューラルネットワークが実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。加えて、又は代替として、試験データセットの各実シーンは、1つ又は複数の空間的に再構成可能な製造ツールを備える実製造シーンである。そのような場合、ドメイン適応型ニューラルネットワークは、実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。

10

【0033】

そのような例では、ドメイン適応型ニューラルネットワークがいずれの場合にも、実製造シーンにおける空間的に再構成可能な製造ツール（例えば、空間的に再構成可能な製造ロボット）の推論のために構成される。ドメイン適応学習方法によってもたらされる前述の改善は、これらの例で特に強調される。特に、空間的に再構成可能な製造ツールを含む実製造シーンの大きな訓練セットに頼ることは困難であるが、その理由はそのようなセットが典型的には機密及び/又は公開であるが、あまりアノテーションが付けられていないからである。他方、空間的に再構成可能な製造ツールを含む仮想製造シーンの大きなデータセットを訓練セットとして提供することは、そのようなデータセットが例えば、仮想製造環境において仮想シミュレーションを生成することができるソフトウェアを使用することによって、容易に取得され、自動的にアノテーション付けされ得るので、特に便利である。

20

【0034】

本発明はまた、ドメイン適応学習方法に従って学習可能なドメイン適応ニューラルネットワークを提供する。

30

【0035】

本発明はまた、機械学習の第2のコンピュータ実施方法を提供する。第2の方法は、シーンのテストデータセットを提供することを含む。テストデータセットは、テストドメインに属する。第2の方法は、ドメイン適応ニューラルネットワークを提供することを含む。ドメイン適応ニューラルネットワークは、訓練ドメインから得られたデータについて学習された機械学習ニューラルネットワークである。ドメイン適応ニューラルネットワークは、テストドメインのシーンにおける空間的に再構成可能なオブジェクトの推論のために構成される。第2の方法は、中間ドメインを決定することをさらに含む。中間ドメインはデータ分布の観点から、テストドメインよりも訓練ドメインに近い。第2の方法は、ドメイン適応ニューラルネットワークを適用することによって、中間ドメイン上で転送されたテストドメインのシーンから空間的に再構成可能なオブジェクトを推論することをさらに含む。この第2の方法は、「ドメイン適応推論法」と呼ばれることがある。

40

【0036】

ドメイン適応推論方法は、空間的に再構成可能なオブジェクトを含むシーンのデータセットを用いて機械学習の改善された方法を形成する。

【0037】

特に、ドメイン適応推論方法は、シーンから空間的に構成可能なオブジェクトの推論を可能にする。これにより、製造シーンにおける空間的に再構成可能な製造ツール（例えば、産業用多関節ロボット）の推論のプロセスのような、空間的に再構成可能なオブジェクトの推論のプロセスにおいて、ドメイン適応推論方法を使用することができる。

50

【 0 0 3 8 】

さらに、この方法はテストデータセットのシーンから空間的に再構成可能なオブジェクトを正確に推論することを可能にし、その一方で、そのような推論に使用されるドメイン適応ニューラルネットワークは訓練ドメインから得られたデータに基づいて機械学習され、訓練ドメインはデータ分布に関してテストドメイン（テストデータセットが属する）から離れている可能性がある。このような改善は特に、中間ドメインの決定によって達成される。実際、ドメイン適応ニューラルネットワークをより良く使用するために、ドメイン適応推論方法は推論のためにドメイン適応ニューラルネットワークを適用する前に、テストドメインのシーンを中間ドメインに属するシーン、換言すれば、データ分布に関してドメイン適応ニューラルネットワークの学習シーンにより近いシーンに変換する。言い換えれば、ドメイン適応推論方法はドメイン適応ニューラルネットワークの出力（それぞれ、テストドメインのそれぞれのシーンにおいて推論される 1 つ又は複数の空間的に再構成可能なオブジェクトである）の精度及び／又は品質を改善するために、ドメイン適応ニューラルネットワークへの入力として、テストドメインのシーンではなく、ドメイン適応ニューラルネットワークが学習されたシーンにより近いシーンを提供することを可能にする。

10

【 0 0 3 9 】

したがって、ドメイン適応推論方法によってもたらされる改善は、テストドメイン及び訓練ドメインが互いに比較的離れている例において特に強調される。これは、例えば、訓練ドメインが仮想シーンで作られ、テストドメインが実シーンで作られている場合である。これは例えば、訓練ドメイン及びテストドメインの両方が仮想シーンで作られているが、テストドメインのシーンが訓練ドメインのシーンよりもフォトリアリスティックである場合にも当てはまる。これらの全ての例において、本方法は（例えば、ドメイン適応ニューラルネットワークのテスト段階において）ドメイン適応ニューラルネットワークへの入力として提供されるテストドメインのシーンの前処理の段階、すなわち、中間ドメインの決定を含み、（データ分布に関して定量化されてもよい）フォトリアリズムに関して訓練ドメインのシーンにより近づける。その結果、ドメイン適応ニューラルネットワークへの入力として提供されるシーンのフォトリアリズムのレベルは、ドメイン適応ニューラルネットワークが学習されたシーンのフォトリアリズムに近づく。これにより、ドメイン適応型ニューラルネットワークの出力（すなわち、テストドメインのシーン上で推論された空間的に再構成可能なオブジェクト）のより良好な品質が保証される。

20

30

【 0 0 4 0 】

実施例では、訓練ドメインが各々 1 つ以上の空間的に再構成可能なオブジェクトを含む仮想シーンの訓練データセットを含む。さらに、仮想シーンのデータセットの各仮想シーンは 1 つ又は複数の空間的に再構成可能な製造ツールを含む仮想製造シーンであってもよく、このような場合、ドメイン適応ニューラルネットワークは実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。これに加えて、又はこれに代えて、テストデータセットは、実シーンを含む。そのような場合の例では試験データセットの各実シーンが 1 つ又は複数の空間的に再構成可能な製造ツールを含む実製造シーンであり、ドメイン適応型ニューラルネットワークは実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。

40

【 0 0 4 1 】

これらの例では、ドメイン適応型ニューラルネットワークがいずれの場合にも、実製造シーンにおける空間的に再構成可能な製造ツール（例えば、空間的に再構成可能な製造ロボット）の推論のために構成され得る。ドメイン適応推論方法によってもたらされる前述の改善は、これらの例で特に強調される。特に、空間的に再構成可能な製造ツールを含む実製造シーンの大きな訓練セットに頼ることは困難であるが、その理由はそのようなセットが典型的には機密及び／又は公開であるが、あまりアノテーションが付けられていないからである。一方、訓練セットとして、空間的に再構成可能な製造ツールを含む仮想製造シーンの大きなデータセットを使用することは、そのようなデータセットが例えば、仮想製造環境において仮想シミュレーションを生成することができるソフトウェアを使用する

50

ことによって、容易に取得され、自動的にアノテーション付けされ得るので、特に便利である。少なくともこれらの理由のために、ドメイン適応ニューラルネットワークは、仮想製造シーンの訓練セット上で学習済である可能性がある。したがって、中間ドメインの決定によって実行されるテストデータセットの前述の前処理はドメイン適応ニューラルネットワークのより良い出力品質を保証するために、データ分布に関して訓練セットのシーンにより近いテストデータセットの実製造シーンを賢明にもたらすことによって、ドメイン適応ニューラルネットワークの訓練セットを考慮に入れる。

【 0 0 4 2 】

ドメイン適応推論法とドメイン適応学習法は独立に行うことができる。特に、ドメイン適応推論方法に従って提供されるドメイン適応ニューラルネットワークは、ドメイン適応学習方法に従って、又は任意の他の機械学習方法によって学習されていてもよい。ドメイン適応推論方法に従ってドメイン適応ニューラルネットワークを提供することは、（例えば、ドメイン適応学習方法に従って）その学習の後にドメイン適応ニューラルネットワークが記憶されているデータ記憶媒体に、例えば、ネットワークを介して遠隔でアクセスすることと、データベースからドメイン適応ニューラルネットワークを検索することを含むことができる。

10

【 0 0 4 3 】

例では、訓練ドメインから取得されたデータが別の中間ドメインのシーンを含む。これらの例では、ドメイン適応ニューラルネットワークが別の中間ドメインで学習済である。別の中間ドメインは、データ分布に関して訓練ドメインよりも中間ドメインに近い。このような例では、ドメイン適応型ニューラルネットワークがドメイン適応型学習方法に従って学習可能なドメイン適応型ニューラルネットワークであってもよい。ドメイン適応学習方法によるテストデータセットはドメイン適応推論方法によるテストデータセットと等しくてもよく、ドメイン適応推論方法によるテストドメインはドメイン適応学習方法による第2のドメインと等しくてもよい。ドメイン適応推論方法による訓練ドメインは、ドメイン適応学習方法による第1のドメインに等しくてもよく、訓練ドメインから取得されたデータはドメイン適応学習方法によって決定された第3のドメインに属してもよい（例えば、形成してもよい）。別の中間ドメインは、ドメイン適応学習方法に従って決定された第3のドメインであってもよい。

20

【 0 0 4 4 】

このような例はドメイン適応学習方法及びドメイン適応推論方法によってもたらされる前述の改善を組み合わせる。すなわち、ドメイン適応ニューラルネットワークの訓練セットとして正しくアノテーション付けされた仮想シーンの大きなデータセットの使用、ドメイン適応ニューラルネットワークの出力の品質及び/又は精度を改善する第3のドメイン（又は別の中間ドメイン）の決定によって現行される訓練セットの前処理段階、及び、ドメイン適応ニューラルネットワークの出力の品質及び/又は精度をさらに改善する中間ドメインの決定によって現行される現実シーンのテストデータセットの前処理段階である。

30

【 0 0 4 5 】

ドメイン適応学習方法及びドメイン適応推論方法は代替的に、同じコンピュータ実装プロセスに統合されてもよい。図1は、ここで説明されるプロセスを示すフローチャートを示す。

40

【 0 0 4 6 】

このプロセスは、ドメイン適応学習方法を統合するオフライン段階を含む。オフライン段階は、ドメイン適応学習方法に従って、仮想シーンのデータセット及び現実シーンのテストデータセットを提供するステップS10を含む。仮想シーンのデータセットは第1のドメインに属し、第1のドメインは、（理論的には前述のように）ドメイン適応学習方法に従って学習されたドメイン適応ニューラルネットワークの訓練ドメインを形成する。現実シーンのテストデータセットは第2のドメインに属し、第2のドメインは、ドメイン適応学習方法に従って学習されたドメイン適応ニューラルネットワークのテストドメインを

50

試験データセットの前述の前処理を含み、それによってドメイン適応ニューラルネットワークの出力の品質を改善する。その結果、オンライン段階での実製造シーンにおける空間的に再構成可能な製造ツールを推定するステップS 6 0は、特に正確かつロバストである。したがって、このプロセスは、実製造シーンにおける空間的に再構成可能な製造ツールの推論の特にロバストで正確なプロセスを構成することができる。

【 0 0 5 1 】

ドメイン適応学習法、ドメイン適応推論法、及びプロセスは、コンピュータ実装される。次に、コンピュータ実施方法（又は処理）の概念について説明する。

【 0 0 5 2 】

「方法（又は処理）がコンピュータによって実施される」とは方法（又は処理）のステップ（又は実質的にすべてのステップ）が少なくとも1つのコンピュータ、又は任意の同様のシステムによって実行されることを意味する。したがって、本方法のステップ（又は処理）はコンピュータによって、場合によっては完全に自動的に、又は半自動的に実行される。例では、方法（又は処理）のステップのうちの少なくともいくつかのトリガがユーザ/コンピュータ対話を介して実行され得る。必要とされるユーザ/コンピュータ対話のレベルは予測される自動化のレベルに依存し、ユーザの希望を実施する必要性とバランスをとることができる。例では、このレベルがユーザ定義及び/又は事前定義され得る。

【 0 0 5 3 】

方法（又はプロセス）のコンピュータ実装の典型的な例は、この目的のために適合されたシステムを用いて方法（又はプロセス）を実行することである。システムはメモリに結合されたプロセッサと、グラフィカルユーザインターフェース（G U I）とを含み、メモリは、本方法を実行するための命令を含むコンピュータプログラムに記録されたものである。メモリはまた、データベースを記憶してもよい。メモリはそのような記憶装置に適合された任意のハードウェアであり、場合によっては、いくつかの物理的に別個の部分（例えば、プログラムのための部分、及び場合によってはデータベースのための部分）を備える。

【 0 0 5 4 】

図5はシステムのG U Iの一例を示し、システムはC A Dシステムである。

【 0 0 5 5 】

G U I 2 1 0 0は、標準的なメニューバー2 1 1 0、2 1 2 0、ならびに底部及び側部ツールバー2 1 4 0、2 1 5 0を有する、典型的なC A D様インターフェースであってもよい。このようなメニューバー及びツールバーはユーザが選択可能なアイコンのセットを含み、各アイコンは当技術分野で知られているように、1つ又は複数の操作又は関数に関連付けられている。これらのアイコンのいくつかは、G U I 2 1 0 0に表示された3 Dモデル化オブジェクト2 0 0 0を編集及び/又は作業するように適合されたソフトウェアツールに関連付けられる。ソフトウェアツールは、ワークベンチにグループ化することができる。各ワークベンチは、ソフトウェアツールの部分集合を含む。特に、ワークベンチの1つは、モデル化された製品2 0 0 0の幾何学的特徴を編集するのに適した編集ワークベンチである。動作中、設計者は例えば、オブジェクト2 0 0 0の一部を事前に選択し、次いで、適切なアイコンを選択することによって、動作（例えば、寸法、色などを変更する）又は幾何学的制約を編集することができる。例えば、典型的なC A D動作は、画面上に表示される3 Dモデル化オブジェクトの打ち抜き加工又は折り畳みのモデル化である。G U Iは例えば、表示された製品2 0 0 0に関連するデータ2 5 0 0を表示することができる。図の例では「特徴木」として表示されるデータ2 5 0 0、及びそれらの3 D表現2 0 0 0はブレーキキャリパ及びディスクを含むブレーキアセンブリに関する。G U Iは編集された製品の動作のシミュレーションをトリガするために、又は表示された製品2 0 0 0の様々な属性をレンダリングするために、例えば、オブジェクトの3 D配向を容易にするための様々なタイプのグラフィックツール2 1 3 0、2 0 7 0、2 0 8 0をさらに示すことができる。カーソル2 0 6 0はユーザがグラフィックツールと対話することを可能にするために、触覚デバイスによって制御され得る。

10

20

30

40

50

【 0 0 5 6 】

図 6 は、システムがクライアントコンピュータシステム、例えばユーザのワークステーションであるシステムの一例を示す。

【 0 0 5 7 】

この例のクライアントコンピュータは、内部通信バス 1 0 0 0 に接続された中央処理装置 (C P U) 1 0 1 0 と、やはりバスに接続されたランダムアクセスメモリ (R A M) 1 0 7 0 とを備える。クライアントコンピュータには、さらに、 B U S に接続されたビデオランダムアクセスメモリ 1 1 0 0 に関連するグラフィカルプロセッシングユニット (G P U) 1 1 1 0 が設けられている。ビデオ R A M 1 1 0 0 は、当技術分野ではフレームバッファとしても知られている。大容量記憶装置制御装置 1 0 2 0 は、ハードドライブ 1 0 3 0 などの大容量記憶装置へのアクセスを管理する。コンピュータプログラム命令及びデータを有形に具現化するのに適した大容量メモリデバイスは、例として、 E P R O M 、 E E P R O M 、 及びフラッシュメモリデバイスなどの半導体メモリデバイス、内部ハードディスク及びリムーバブルディスクなどの磁気ディスク、光磁気ディスク、ならびに C D - R O M ディスク 1 0 4 0 を含む、すべての形態の不揮発性メモリを含む。前述のいずれも、特別に設計された A S I C (特定用途向け集積回路) によって補足されるか、又はその中に組み込まれてもよい。ネットワークアダプタ 1 0 5 0 は、ネットワーク 1 0 6 0 へのアクセスを管理する。クライアントコンピュータは、カーソル制御デバイス、キーボードなどの触覚デバイス 1 0 9 0 も含むことができる。ユーザがディスプレイ 1 0 8 0 上の任意の所望の位置にカーソルを選択的に位置決めすることを可能にするために、カーソル制御装置がクライアントコンピュータ内で使用される。さらに、カーソル制御装置はユーザが様々なコマンドを選択し、制御信号を入力することを可能にする。カーソル制御装置は、システムに制御信号を入力するための多数の信号発生装置を含む。典型的にはカーソル制御装置がマウスであってもよく、マウスのボタンは信号を生成するために使用される。代替的に又は追加的に、クライアントコンピュータシステムは、センシティブパッド及び / 又はセンシティブスクリーンを含むことができる。

【 0 0 5 8 】

コンピュータプログラムはコンピュータによって実行可能な命令を含むことができ、命令は上記のシステムに、ドメイン適応学習方法、ドメイン適応推論方法、及び / 又は処理を実行させるための手段を含む。プログラムは、システムのメモリを含む任意のデータ記憶媒体に記録可能であってもよい。プログラムは例えば、デジタル電子回路において、又はコンピュータハードウェア、ファームウェア、ソフトウェアにおいて、又はそれらの組み合わせにおいて実装されてもよい。プログラムは、装置、例えば、プログラマブルプロセッサによる実行のための機械可読記憶デバイスに有形に具現化された製品として実装されてもよい。プロセス / 方法ステップ (すなわち、ドメイン適応学習方法、ドメイン適応推論方法、及び / 又はプロセスのステップ) は入力データに対して動作し、出力を生成することによって、プロセスの機能を実行するための命令のプログラムを実行するプログラマブルプロセッサによって実行され得る。したがって、プロセッサはプログラム可能であり、データ記憶システム、少なくとも 1 つの入力デバイス、及び少なくとも 1 つの出力デバイスからデータ及び命令を受信し、それらにデータ及び命令を送信するように結合され得る。アプリケーションプログラムは、高レベルの手続き型又はオブジェクト指向プログラミング言語で、あるいは必要に応じてアセンブリ言語又は機械語で実装することができる。いずれの場合も、言語は、コンパイルされた言語又は解釈された言語であってもよい。プログラムはフルインストールプログラムであってもよいし、更新プログラムであってもよい。システムへのプログラムのアプリケーションは、いずれにしても、ドメイン適応学習方法、ドメイン適応推論方法、及び / 又は処理を実行するための命令に帰着する。

【 0 0 5 9 】

ドメイン適応学習法、ドメイン適応推論法、及びプロセスに含まれるシーンのデータセットを提供する概念を説明する。この概念の議論の前に、そこに含まれるデータ構造がここで議論される。理解されるように、本明細書で提供されるデータ構造定義及び例はドメ

10

20

30

40

50

イン適応学習方法、ドメイン適応推論方法、及び／又は処理を提供する任意のデータセットの少なくとも一部（たとえば、すべて）に適用することができる。

【0060】

本開示のコンテキストでは、シーンがある時点における背景内の１つ又は複数のオブジェクトの配置（例えば、配備）を指定する。バックグラウンドは典型的には実世界の物理的環境を表し、オブジェクトは典型的には実世界の物理的オブジェクトを表し、オブジェクトの配置は、典型的には特定時点における実世界環境における現実世界の物体の配置を表す。これは、シーンがそれらの実世界環境における現実世界の物体のある時点での表現であるということに等しい。この表現は、実在、幾何学的に現実的、意味論的に現実的、写真現実的、及び／又は仮想的であってもよい。実際の、幾何学的に現実的な、意味論的に現実的な、写真的に現実的な、及び仮想シーンの概念を以下に説明する。シーンは、典型的には（常にではないが）３次元である。以下において、シーン（例えば、仮想又は現実）とその表現（例えば、仮想又は現実画像による）との間の区別は、簡略化のために行われない。

10

【0061】

本開示のコンテキストでは、シーンのデータセットが典型的には多数のシーン、例えば、１０００、１００００、１０００００、又は１００００００を超えるシーンを含む。本開示のシーンの任意のデータセットは同じコンテキストからのシーン（例えば、製造シーン、建設現場シーン、ハーバーシーン、又はターマックシーン）から作成されてもよく、又は実質的に作成されてもよい。ドメイン適応学習方法、ドメイン適応推論方法、及び／又は処理に含まれるシーンは実際にはすべて同じコンテキストを形成するシーン（例えば、製造シーン、建設現場シーン、ハーバーシーン、又はターマックシーン）であってもよい。

20

【0062】

ドメイン適応学習法、ドメイン適応推論法及び／又は処理に含まれる任意のシーンは、１つ又は複数の空間的再構成可能オブジェクトを含み得る。空間的に再構成可能なオブジェクトはいくつかの（例えば、機械的である）部品を含むオブジェクトであり、オブジェクトは、オブジェクトの前記部品の間に１つ以上の意味的に現実的な空間的關係が存在することを特徴とする。空間的に再構成可能な物体は典型的には互いに物理的に連結された部品の（例えば、機械的である）アセンブリであってもよく、アセンブリは、少なくとも１つの自由度を有する。これは、物体の少なくとも第１の部品が物体の少なくとも第２の部品に対して移動可能であり、一方、少なくとも第１の部品及び少なくとも第２の部品の両方がアセンブリに物理的に取り付けられたままであることを意味する。例えば、少なくとも第１の部品は、少なくとも第２の部品に対して平行移動及び／又は回転されてもよい。したがって、空間的に再構成可能なオブジェクトは非常に多数の空間的構成及び／又は位置を有する可能性があり、それによって、適切な技能を有していない人物のために、オブジェクトを識別し、たとえばラベル付けすることが困難になる。本開示のコンテキストでは、空間的に再構成可能な物体が製造シーンにおける空間的に再構成可能な製造ツール（産業用多関節ロボットなど）、建設現場シーンにおけるクレーン、ターマックシーンにおける空中通路、又は港湾シーンにおけるポートクレーンとすることができる。

30

40

【0063】

本開示のコンテキストでは、シーンは製造シーンであってもよい。製造シーンは、現実世界の製造環境における現実世界の物体を表すシーンである。製造環境は、工場又は工場の一部であってもよい。工場又は工場の一部の物体は製造された製品、すなわち、現場に関連する時点で工場において実行される１つ又は複数の製造工程によって既に製造された製品とすることができる。工場又は工場の一部の物体は製造される製品、すなわち、シーンに関連する時点で工場において実行される１つ又は複数の製造工程によって製造される製品であってもよい。工場内の物体又は工場の一部はまた、（例えば、空間的に再構成可能である）製造ツールであってもよく、これは、工場で実行される１つ以上の製造工程に参与するツールである。また、工場内のオブジェクトは、工場の背景を構成するオブジェ

50

クトであってもよい。

【 0 0 6 4 】

言い換えれば、工場又は工場の一部を表すシーンはしたがって、それぞれが工場の1つ又は複数の製造工程によって製造されている、又は既に製造されている1つ又は複数の製品と、たとえば、それぞれが前記製品の1つ又は複数の製造に関与する1つ又は複数の製造ツールとを含むことができる。製品は典型的には部品の（例えば、機械的である）部品又はアセンブリであってもよい（又は、部品のアセンブリが本開示の観点からそれ自体が部品として見られてもよいので、同等に部品のアセンブリであってもよい）。

【 0 0 6 5 】

機械部品は陸上車両（例えば、自動車及び軽トラック機器、レーシングカー、オートバイ、トラック及びモータ機器、トラック及びバス、列車を含む）の一部、航空機の一部（例えば、機体機器、航空宇宙機器、推進機器、防衛機器、航空機機器、宇宙機器を含む）、海軍車両の一部（例えば、海軍機器、商用船舶、オフショア機器、ヨット及び作業船を含む）、一般機械部品（例えば、工業製造機械、重移動機械又は機器、設置機器、工業機器、加工金属製品、タイヤ製造製品、ロボットアームのような関節式及び／又は再構成可能な製造機器を含む）、電気機械又は電子部品（例えば、家庭用電子機器、セキュリティ及び／又は制御及び／又は計装製品、コンピュータ及び通信機器、半導体、医療デバイス及び機器を含む）、消費財（例えば、家具、家庭用品、庭用品、レジャー用品、ファッション用品、ハードグッズ小売業者の製品、ソフトグッズ小売業者の製品、包装（例えば、食品及び飲料及びタバコ、美容及びパーソナルケア、家庭用品包装を含む）を含む）であってもよい。

【 0 0 6 6 】

機械部品はまた、成形部品（すなわち、成形製造工程によって製造された部品）、機械加工部品（すなわち、機械加工製造工程によって製造された部品）、穿孔部品（すなわち、穿孔製造工程によって製造された部品）、旋削部品（すなわち、旋削製造工程によって製造された部品）、鍛造部品（すなわち、鍛造製造工程によって製造された部品）、打ち抜き部品（すなわち、打ち抜き製造工程によって製造された部品）、及び／又は折り曲げ部品（すなわち、折り曲げ製造工程によって製造された部品）のうちの1つ又は妥当な組み合わせとすることができる。

【 0 0 6 7 】

製造ツールは、以下のいずれかであってもよい。

・ブローチ盤、ドリルプレス、ギヤシェーパ、ホブ盤、ホーン、旋盤、旋盤、フライス盤、板金せん断力などの加工ツール（すなわち、加工プロセスの少なくとも一部を実施するツール）、シェーパ、鋸、プレーナ、スチュワートプラットフォームミル、研削機、複合加工機（例えば、旋削、フライス加工、研削、及び／又は材料取り扱いを1つの高度に自動化された機械ツールに結合する多くの軸を有する）などの加工ツール（すなわち、加工プロセスの少なくとも一部を実施するツール）。

・圧縮成形機（すなわち、圧縮成形プロセスの少なくとも一部を実行する機械）であって、典型的には、フラッシュブランジャー種類、ストレートブランジャー種類、又はランドブランジャー種類金種類などの少なくとも1つの金種類又は成形行列を含む。

・ダイカスト機、金属射出成形機、プラスチック射出成形機、液体シリコン射出成形機又は反応射出成形機のような射出成形機（すなわち、射出成形プロセスの少なくとも一部を実行する機械）。

・ドリルビット、皿穴、座ぐり、タップ、ダイ、フライス、リーマ、又はコールドソーブレードなどの回転切削工具（例えば、ドリル加工の少なくとも一部を実行する）。

・非回転切削工具（例えば、旋削プロセスの少なくとも一部を実行する）、例えば、先端付き工具又は成形工具。

・機械的鍛造プレス（典型的には少なくとも1つの鍛造ダイを含む）、油圧鍛造プレス（典型的には少なくとも1つの鍛造ダイを含む）、圧縮空気によって動力を供給される鍛造ハンマー、電気によって動力を供給される鍛造ハンマー、油圧によって動力を供給され

10

20

30

40

50

る鍛造ハンマー、又は蒸気によって動力を供給される鍛造ハンマーなどの鍛造機械（すなわち鍛造プロセスの少なくとも一部を実行する）。

・スタンピングプレス、プレス、パンチング機械、ブランキング機械、エンボス加工機械、曲げ加工機械、フランジング機械、又はコイニング機械などのスタンピング機械（すなわち、スタンピングプロセスの少なくとも一部を実行する）。

・曲げ又は折り曲げ機械（すなわち、曲げ又は折り曲げ処理の少なくとも一部を実行する）、そのようなボックス及びパンブレーキ、プレスブレーキ、フォルダ、パネルベンダ又は機械プレス。

・スポット溶接ロボット、又は（例えば電気）自動車塗装ロボットのような電気塗装ツール（即ち、製品の塗装プロセスの少なくとも一部を実行する）。

10

【 0 0 6 8 】

上に列挙された製造ツールのうちの任意の 1 つは、空間的に再構成可能な製造ツール、すなわち、空間的に再構成可能な製造ツールであってもよく、又は空間的に再構成可能な製造ツールの少なくとも一部を形成してもよい。例えば、空間的に再構成可能な製造ツールのツールホルダによって保持される製造ツールは、空間的に再構成可能な製造ツールの少なくとも一部を形成する。空間的に再構成可能な製造ツール（例えば、上記に列挙されたツールのいずれか 1 つ）は（例えば、関節式）産業用ロボットのような製造ロボットであってもよい。

【 0 0 6 9 】

次に、現実のシーンの概念について説明する。前述のように、シーンは現実世界環境（すなわち、シーンの背景）における現実世界の物体の配置を表す。シーンは、配置の表現が現実世界におけるものと同じ物理的配置の獲得に由来する物理的配置である場合に現実である。取得はデジタル画像取得プロセスによって実行されてもよく、デジタル画像取得プロセスは環境内の物体の物理的配置の視覚的特徴のデジタル的に符号化された表現を生成する。このようなデジタル的に符号化された表現は簡潔にするために、以下では「画像」又は「デジタル画像」と呼ぶことがある。デジタル画像取得プロセスは、任意選択で、そのような画像の処理、圧縮、及び／又は記憶を含むことができる。デジタル画像は、カメラ又は同様の装置、スキャン又は同様の装置、及び／又はビデオカメラ又は同様の装置によって、物理的環境から直接作成されてもよい。あるいは、デジタル画像が画像スキャナ又は同様の装置によって、写真、写真フィルム（例えば、デジタル画像はフィルムの瞬間のスナップショットである）、又は印刷された紙などの同様の媒体内の別の画像から取得されてもよい。デジタル画像はまた、断層撮影装置、スキャン、サイドスキャンソナー、ラジオ望遠鏡、X 線検出器及び／又は光刺激性蛍光板（PSP）を用いて取得されたもののなどの非画像データの処理によって取得されてもよい。本開示のコンテキストでは、実製造シーンが例えば、工場内に配置されたカメラ（又は同様の装置）によって得られたデジタル画像に由来してもよい。代替的に、製造シーンは、工場に配置された 1 つ以上のビデオカメラによって取得されたフィルムのスナップショットに由来してもよい。

20

30

【 0 0 7 0 】

図 7 ~ 図 12 は実生産シーンの例を示しており、各シーンは、関節式産業用ロボット 70、72、80、90、92、94、96、98、100、102、104、110、112、120、122、124 から構成されている。

40

【 0 0 7 1 】

次に、仮想シーンの概念について説明する。シーンは、コンピュータによって生成されたデータに由来する場合、仮想的である。そのようなデータは（例えば、3D）シミュレーションソフトウェア、（例えば、3D）モデリングソフトウェア、CADソフトウェア、又はビデオゲームによって生成され得る。これは、シーン自体がコンピュータ生成されているということである。仮想シーンのオブジェクトは、仮想シーン内の現実的な位置に配置されてもよい。あるいは、仮想シーンのオブジェクトが例えば、特に後述するように、前記オブジェクトが既存の仮想シーンにランダムに配置されるときに、仮想シーン内の非現実的な位置に配置されてもよい。これは、仮想シーンが仮想背景（例えば、現実世界

50

環境の仮想表現)における仮想物体(例えば、現実世界の物体の仮想表現)の配置を表すことを意味する。前記配置は仮想物体の少なくとも一部に対して物理的に現実的であってもよく、すなわち、前記少なくとも一部の仮想物体の配置は、現実世界の物体の現実世界の物理的配置を表す。追加的に又は代替的に、仮想物体の少なくとも別の部分は非現実的な配置(すなわち、現実世界の物理的配置に対応しない)で(例えば、互いに対して及び/又は前記少なくとも部分のオブジェクトに対して)配置されてもよく、例えば、前記少なくとも部分の仮想物体は特に後述されるように、ランダムに配置又は配置される。仮想シーンの任意のオブジェクトに、自動的にアノテーションを付けたり、ラベルを付けることができる。オブジェクトをラベル付けする概念については、以下でさらに説明する。仮想製造シーンは、製造環境のための3Dモデリングソフトウェア、製造環境のための3Dシミュレーションソフトウェア、又はCADソフトウェアによって生成され得る。仮想製造シーン内の任意のオブジェクト(製品又は製造ツールなど)は立体(例えば、B-repモデル)のスキン(例えば、外面)を表す3Dモデル化オブジェクトによって仮想的に表すことができる。3Dモデル化オブジェクトは、CADシステムを有するユーザによって設計されていてもよく、及び/又は製造環境のための3Dシミュレーションソフトウェアによって実行される仮想製造シミュレーションに由来していてもよい。

【0072】

モデル化されたオブジェクトは、例えばデータベースに格納されたデータによって定義される任意のオブジェクトである。伸長により、表現「モデル化オブジェクト」は、データ自体を指定する。ドメイン適応学習方法、ドメイン適応推論方法、及び/又は処理を実行するシステムのタイプに応じて、モデル化されたオブジェクトは、異なる種類のデータによって定義されてもよい。システムは、実際にはCADシステム、CAEシステム、CAMシステム、PDMシステム、及び/又はPLMシステムの任意の組み合わせであってもよい。これらの異なるシステムでは、モデル化されたオブジェクトが対応するデータによって定義される。したがって、CADオブジェクト、PLMオブジェクト、PDMオブジェクト、CAEオブジェクト、CAMオブジェクト、CADデータ、PLMデータ、PDMデータ、CAMデータ、CAEデータについて言うことができる。しかしながら、モデル化オブジェクトはこれらのシステムの任意の組み合わせに対応するデータによって定義され得るので、これらのシステムは他のシステムのうちの1つを排他的にするものではない。したがって、システムは以下に提供されるそのようなシステムの定義から明らかになるように、CADシステム及びPLMシステムの両方であってもよい。

【0073】

CADシステムとは、さらに、CATIAのような、モデル化オブジェクトのグラフィック表現に基づいてモデル化オブジェクトを少なくとも設計するように適合された任意のシステムを意味する。この場合、モデル化オブジェクトを定義するデータは、モデル化オブジェクトの表現を可能にするデータを含む。CADシステムは例えば、ある場合には、面又は表面を有するエッジ又は線を使用してCADモデル化オブジェクトの表現を提供することができる。線、エッジ、又は表面は様々な方法、例えば、不均一有理Bスプライン(NURBS)で表すことができる。具体的には、CADファイルは仕様を含み、そこから幾何学的形状を生成することができ、これにより表現を生成することができる。モデル化されたオブジェクトの仕様は、単一のCADファイル又は複数のCADファイルに格納することができる。CADシステム内のモデル化オブジェクトを表すファイルの典型的なサイズは、部品当たり1メガバイトの範囲内である。そして、モデル化オブジェクトは、典型的には何千もの部品のアセンブリであってもよい。

【0074】

CAMソリューションとは、さらに、製品の製造データを管理するように適合された任意のソリューション、ハードウェアのソフトウェアを意味する。製造データは一般に、製造する製品、製造工程、及び必要な資源に関するデータを含む。CAMソリューションは製品の製造処理全体を計画し、最適化するために使用される。例えば、CAMユーザに、製造処理の実現可能性、製造処理の継続時間、又は製造処理の特定のステップで使用する

10

20

30

40

50

ことができる特定のロボットなどのリソースの数に関する情報を提供することができ、したがって、管理又は必要な投資に関する決定を可能にする。CAMは、CAD処理及び潜在的CAE処理の後の後続処理である。このようなCAMソリューションは、ダッソーシステムズによってDELMIA（登録商標）の商標で提供されている。

【0075】

CADのコンテキストでは、モデル化されたオブジェクトが典型的には例えば、部品又は部品のアセンブリ、又は場合によっては製品のアセンブリなどの製品を表す3Dモデル化されたオブジェクトとすることができる。「3Dモデリングされたオブジェクト」とは、その3D表現を可能にするデータによってモデリングされる任意のオブジェクトを意味する。3D表現は、全ての角度から部品を見ることを可能にする。例えば、3Dモデル化オブジェクトは3D表現される場合、その軸のいずれかの周り、又は表現が表示される画面内のいずれかの軸の周りで取り扱われ、回転され得る。これは特に、3Dモデル化されていない2Dアイコンを除外する。3D表現の表示は設計を容易にする（すなわち、設計者が統計的に彼らのタスクを達成する速度を増加させる）。これは、製品の設計が製造工程の一部であるため、産業界における製造工程を高速化する。

【0076】

図13～図17は、仮想多関節産業用ロボット130、140、150、152、160、162、170、172をそれぞれ備える仮想製造場面を示す。

【0077】

仮想シーンは幾何学的及び意味論的に現実的であってもよく、これは、シーン内の仮想物体（製造シーン内のプロダクト又は製造工具など）の幾何学及び機能性が現実的で例えば、現実世界におけるように対応する幾何学及び機能性を現実的に表すことを手段する。仮想シーンは、フォトリアリスティックではなく、幾何学的及び意味的にリアリスティックであってもよい。代替的に、仮想シーンは幾何学的及び意味的に現実的であり、フォトリアリスティックであってもよい。

【0078】

フォトリアルなイメージの概念はシェーディング（つまり、表面の色と明るさがライティングによってどのように変化するか）、テクスチャマッピング（つまり、表面に細部を適用する方法）、バンプマッピング（つまり、表面に小さなスケールの凹凸をシミュレートする方法）、フォギング/参加媒体（つまり、明瞭でない大気や空気を通過するときどのように光が暗くなるか）、影（つまり、部分的に隠れた光源によって生じる暗さが変化するか）、反射（つまり、鏡面のような、非常に光沢のある反射）、透明度又は不透明度（つまり、立体オブジェクトを通過する光の鋭い透過）、屈折（つまり、透明度に関連する光の屈折）、回折（つまり）など、多くの可視的な特徴によって理解できる。曲げ、光線を拡散させ、光線を乱す物体又はアパーチャによって通過する光の干渉、間接照明（つまり、光源から直接照射されるのではなく、他の表面から反射された光によって照射される光の干渉）、焼灼（つまり、間接照明の形式、輝く物体からの光の反射、及び/又は透明な物体を通しての光の焦点合わせ）、明るいハイライトを別の物体に生成する、被写界深度（つまり、物体が焦点を結んでいる物体の前方又は後方に遠すぎる場合に、ぼやけて見える、又は焦点が合っていない）、及び/又はモーションブラー（つまり、物体が高速の動きによってぼやけて見える、又はカメラの動き）。そのような特性の存在はそのようなデータ（すなわち、仮想フォトリアリスティックシーンに対応する、又はそれから生じる）を、実データ（実シーンを表す画像から生じる）と同様の分布を有するようにする。このコンテキストでは、仮想集合分布が現実世界データに近いほど、それがより写実的に見える。

【0079】

ここで、データ配信の概念について説明する。機械学習（ML）モデルを構築する状況では、多くの場合、モデルを訓練し、同じ分布に由来するデータに対してそれをテストすることが望ましい。確率論と統計学では、確率分布が実験で起こりうるさまざまな結果の発生確率を提供する数学関数である。言い換えれば、確率分布は、事象の確率に関するラ

10

20

30

40

50

ンダム現象の記述である。確率分布は、観察されているランダム現象の全ての可能な結果の集合で基礎となるサンプル空間に関して定義される。サンプル空間は実数の集合であってもよいし、より高次元のベクトル空間であってもよいし、非数値のリストであってもよい。仮想データに関する訓練及び実データに関する試験の際に、分布シフトの問題に直面することが理解できる。シフト問題に対処するために実画像を用いて仮想画像を修正するとき、ドメイン適応学習方法、ドメイン適応推論方法、及び／又はプロセスは、仮想データの分布を、実データの分布とよりよく整合するように変換することができる。2つの分布がどのくらい良好に整列するかを定量化するために、いくつかのアプローチを使用することができる。例示のために、以下に2つのアプローチの例を示す。

【0080】

第1のアプローチは、ヒストグラムアプローチである。N個の画像ペアを作成する場合を考える。各画像対は第1の生成された仮想画像と、第1の仮想画像と同じシーンを表す実画像に対応する第2の仮想画像とからなり、実画像は、ジェネレータ仮想 - 実ジェネレータを使用することによって得られる。訓練データを修正することの関連性を評価するために、ドメイン適応学習方法、ドメイン適応推論方法、及び／又はプロセスは、仮想画像分布が合成から現実への修正の前後で現実世界データとどの程度良好に整列するかを定量化する。そうするために、ドメイン適応学習方法、ドメイン適応推論方法、及び／又はプロセスは実画像のセットを考慮し、合成から実への変形例の前後の実データの平均画像と仮想データとの間のユークリッド距離を計算することができる。形式的に、例えば、変形例の前後のそれぞれの仮想画像のImageNet分類タスク（インセプションネットワーク、[22]参照）上で訓練された、事前訓練されたネットワークから導出された深層特徴

【数1】

$$D^v = \{v_1, \dots, v_N\} \quad \text{及び}$$

$$D^{v'} = \{v'_1, \dots, v'_N\}$$

を考察する。現実データセットの平均画像から導出された深層特徴を

【数2】

$$m_{real}$$

と表すと、ユークリッド差分は、

【数3】

$$ED^v = \{v_1 - m, \dots, v_N - m\} \quad \text{及び}$$

$$ED^{v'} = \{v'_1 - m, \dots, v'_N - m\}$$

である。実データ分布との整合を評価するために、ドメイン適応学習方法、ドメイン適応推論方法、及び／又はプロセスはヒストグラムを用いて

【数4】

$$ED^v \quad \text{及び} \quad ED^{v'}$$

の分布を比較するヒストグラムは、数値データの分布の正確な表現である。それは連続変数の確率分布の推定値である。図18は、

【数 4】

$$ED^v \text{ 及び } ED^{v'}$$

の間のユークリッド距離の分布のヒストグラムを示す。図 18 によれば、元の虚像

【数 5】

$$ED^v$$

(青)を使用した場合に得られる分布と比較して、表現が修正された虚像

10

【数 6】

$$ED^{v'}$$

(赤)由来である場合に、分布は 0 付近ではるかにクラスタ化される。このヒストグラムは、修正された仮想データがデータ分布に関してルート仮想データよりも現実世界データに近い新しいドメインに転送されると仮定する、テストされた仮説を検証する。フォトリアリズムの観点から、所与の画像が人工画像であるか実際画像であるかを視覚的に推測するように人々に求めるテストを考える。修正された画像は、元の仮想画像よりも視覚的に人間をだましやすいたことが明らかである。これは、上記のヒストグラムによって評価された分布アラインメントに直接リンクされる。

20

【0081】

第2のアプローチは、フレッシュインセプション距離(FID)アプローチである([21]参照)。FIDは、2組の画像のインセプション埋め込みの分布(インセプションの最終層からの活性化、[22]参照)を比較する。これらの分布は、それぞれの平均と共分散によってパラメータ化された多次元ガウス分布としてモデル化される。形式的には、

【数 7】

$$D^A = \{(x_A)\}$$

30

をシーンの第1のデータセット(例えば、現実シーンのデータセット)とし、

【数 8】

$$D^B = \{(x_B)\}$$

を第2のデータセット(例えば、仮想シーンのデータセット)とする。第2のアプローチは例えば、ImageNet分類タスクで訓練された、事前訓練されたネットワークから派生した特徴エクストラクタネットワークを用いて、それぞれ

40

【数 9】

$$x_A' \text{ (複数) 及び } x_B' \text{ (複数)}$$

から深層特徴

【数 10】

$$v^{A'} \text{ (複数) 及び } v^{B'} \text{ (複数)}$$

50

を抽出する（インセプションネットワーク、[2 2] 参照）。次に、第 2 のアプローチは【数 1 1】

$$v^{A'} \text{（複数）}$$

において特徴を集める。これにより、多数の C 次元ベクトルが得られる。ここで、C は特徴次元である。次に、第 2 のアプローチは、これらのベクトルの平均

【数 1 2】

$$\mu_L^A$$

10

及び共分散行列

【数 1 3】

$$\Sigma_L^A$$

を計算することによって、多変量ガウスをこれらの特徴ベクトルに適合させる。同じように、私たちは

【数 1 4】

20

$$\mu_L^B \text{ 及び } \Sigma_L^B$$

を持っている。これらの抽出された特徴の間の F I D は

【数 1 5】

$$\|\mu_L^A - \mu_L^B\|^2 + \text{Tr}(\Sigma_L^A + \Sigma_L^B) - 2\sqrt{\Sigma_L^A \Sigma_L^B}.$$

である。

30

【0 0 8 2】

距離が小さいほど、

【数 1 6】

$$D^A \text{ 及び } D^B$$

の分布は近くなる。フォトリアリズムの観点から、第 1 のデータセットが現実のシーンのデータセットであり、第 2 のデータセットが仮想シーンのデータセットである場合、距離が小さいほど、写真リアリズムのある画像が得られる。この測定は、分布アラインメントに関して、あるドメインから別のドメインへの我々のデータの修正の関連性を評価することを可能にする。

40

【0 0 8 3】

データ分布に関する前の議論から、シーンの 2 つのデータセット間の距離は、前に議論したユークリッド距離又は F I D などの距離によって定量化することができることになる。したがって、本開示のコンテキストでは「シーンの第 1 のデータセット（例えば、ファミリー、例えば、ドメイン）によって、シーンのドメインの概念はシーンの第 3 のデータセット（例えば、ファミリー、例えば、ドメイン）よりもシーンの第 2 のデータセット（例えば、ファミリー、例えば、ドメイン）に近い」とは第 1 のデータセット（それぞれ、ファミリー又はドメイン）と第 2 のデータセット（それぞれ、ファミリー又はドメイン）

50

との間の距離が第3のデータセット（それぞれ、ファミリー又はドメイン）と第2のデータセット（それぞれ、ファミリー又はドメイン）との間の距離よりも小さいことを意味する。第2のデータセットが実シーンのデータセットであり、3番目と1番目の両方のデータセットが仮想シーンのデータセットである場合、1番目のデータセットは3番目のデータセットよりも写真現実的であると言える。

【0084】

ここで、「ドメイン」の概念について説明する。ドメインは、シーンのファミリーの分布を指定する。したがって、ドメインは、同様に分布のファミリーである。データセットのシーンとファミリーのシーンがデータ分布に関して近い場合、シーンのデータセットはドメインに属する。例では、データセットのシーンとファミリーのシーンとの間の距離（例えば、前述のようなユークリッド又はFID）が所定の閾値よりも小さくなるように、所定の閾値が存在することが手段である。シーンのデータセットはそのようなドメインを形成してもよく、その場合、ドメインは「シーンのドメイン」又は「データセットのシーンから作られたドメイン」と称されてもよいことが理解されるべきである。ドメイン適応学習方法による第1のドメインは仮想シーンのドメイン、例えば、各々が幾何学的及び意味論的に現実的であるが、フォトリアリスティックではない仮想シーンのドメインである。ドメイン適応学習法による第2のドメインは、現実のシーンのドメインである。ドメイン適応学習方法によるテストドメインは、現実のシーンのドメインであってもよい。あるいは、テストドメインが例えば幾何学的及び意味的に現実的であり、任意選択的にフォトリアリスティックである仮想シーンのドメインであってもよい。ドメイン適応学習方法による訓練ドメインは例えば、幾何学的及び意味論的に現実的であり、任意選択的に写真現実的である仮想シーンのドメインであってもよい。例では訓練ドメインがテストドメインのシーンよりもフォトリアリスティックでない仮想シーンのドメインであり、例えば、訓練ドメインは仮想シーンのドメインであり、テストドメインは現実シーンのドメインである。

【0085】

次に、シーンのデータセットを提供する概念について説明する。

【0086】

本開示のコンテキストでは、シーンのデータセットを提供することは自動的に、又はユーザによって実行され得る。例えば、ユーザは、データセットが格納されているメモリからデータセットを検索することができる。さらに、ユーザはすでに取得されたデータセットに1つ以上のシーンを追加することによって（例えば、1つずつ）、データセットを完成させることを選択することができる。一つ以上のシーンを追加することは、一つ以上のメモリからそれらを取得し、データセットに含めることを含むことができる。あるいは、ユーザがデータセットのシーンの少なくとも一部（例えば、全て）を作成してもよい。例えば、仮想シーンのデータセットを提供する場合、ユーザは先に論じたように、3Dシミュレーションソフトウェア及び/又は3Dモデリングソフトウェア及び/又はCADソフトウェアを使用することによって、最初にシーンの少なくとも一部（例えば、すべて）を取得することができる。より一般的には、仮想シーンのデータセットを提供することはドメイン適応学習方法、ドメイン適応推論方法及び/又は処理のコンテキストにおいて、例えば、前述のように、3Dシミュレーションソフトウェア（例えば、製造環境のための）及び/又は3Dモデリングソフトウェア及び/又はCADソフトウェアを使用することによって、シーンを取得する（例えば、計算する）ステップによって先行されてもよい。同様に、ドメイン適応学習方法、ドメイン適応推論方法、及び/又は処理のコンテキストにおいて、現実シーンのデータセットを提供することに先立って、前述のように、例えば、前述のように、デジタル画像取得処理を介して、シーンを取得する（例えば、取得する）ステップを行うことができる。

【0087】

ドメイン適応学習方法及び/又は処理の例では、仮想シーンのデータセットの各仮想シーンが1つ又は複数の空間的に再構成可能な製造ツールを含む仮想製造シーンである。こ

10

20

30

40

50

これらの例では、ドメイン適応型ニューラルネットワークが実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。これらの例では、仮想シーンのデータセットの仮想シーンがそれぞれ、シミュレーションから生じ得る。従って、仮想シーンのデータセットを提供するステップS 10は前述したように、例えば製造環境用の3Dシミュレーションソフトウェアを使用してシミュレーションを実施するステップに先行してもよい。例では、シミュレーションが3D体験（すなわち、正確なモデリングを有し、忠実な（例えば、現実適合する）挙動を有する仮想シーン）とすることができ、所与のシーンの異なる構成、例えば、所与のシーンにおけるオブジェクト異なる配置、及び/又は所与のシーンにおけるいくつかの（例えば、空間的に再構成可能である）オブジェクト異なる構成がシミュレートされる。データセットは特に、前記異なる構成に対応する仮想シーンを含むことができる。

10

【0088】

ドメイン適応学習方法及び/又は処理の例では、試験データセットの各実シーンが1つ又は複数の空間的に再構成可能な製造ツールを含む実製造シーンである。これらの例では、ドメイン適応型ニューラルネットワークが実製造シーンにおける空間的に再構成可能な製造ツールの推論のために構成される。これらの例では、試験データセットの実シーンがそれぞれ、デジタル取得プロセスで取得された画像に由来し得る。したがって、仮想シーンのテストデータセットを提供するステップS 10の前に、例えば、1つ又は複数のスキャン及び/又は1つ又は複数のカメラ及び/又は1つ又は複数のビデオカメラを使用することによって、テストデータセットのシーンが由来する工場の画像を取得するステップを行うことができる。

20

【0089】

ドメイン適応推論方法及び/又は処理の例では、テストデータセットは実シーンを含む。これらの例では試験データセットの各現実のシーンが1つ又は複数の空間的に再構成可能な製造ツールを含む現実の製造シーンとすることができ、ドメイン適応型ニューラルネットワークは現実の製造シーンにおける空間的に再構成可能な製造ツールの推論のために任意選択で構成される。したがって、仮想シーンのテストデータセットを提供するステップS 10の前に、例えば、1つ又は複数のスキャン及び/又は1つ又は複数のカメラ及び/又は1つ又は複数のビデオカメラを使用することによって、テストデータセットのシーンが由来する工場の画像を取得するステップを行うことができる。

30

【0090】

次に、第3のドメインを決定するステップS 20について説明する。

【0091】

第3のドメインを決定するステップS 20はデータ分布の観点から、第3のドメインが第1のドメインよりも第2のドメインに近くなるように行われる。このように、第3のドメインを決定するステップS 20は第1のドメインに属する仮想シーンのデータセットの仮想シーンごとに、推論されたすべての別のシーンがデータ分布に関して第1のドメインよりも第2のドメインに近いドメイン、すなわち第3のドメインを形成するように、別のシーンの推論を指定する。前記別のシーンの推論は仮想シーンの前記別のシーンへの変換（例えば、変換）を含むことができる。変換は例えば、画像を1つのドメインから別のドメインに転送することができるジェネレータを使用して、仮想シーンを前記別のシーンに変換することからなってもよい。そのような変換は仮想シーンの1つ又は複数の部分のみに関係し、前記1つ又は複数の部分は、例えばジェネレータを使用して変換され、一方、シーンの残りは変更されないままであることを理解されたい。前記別のシーンは、1つ以上の部分が変換され、残りのシーンは変更されないシーンから構成されてもよい。あるいは、変換がシーンの残りを（例えば、第2のドメインに属する）他のシーンの一部によって置き換えることを含んでもよい。そのような場合、前記別のシーンは、1つ又は複数の部分が変換され、残りのシーンが何らかの他のシーンの前記部分によって置き換えられているシーン内であってもよい。推論は第1のドメインのすべてのシーンについて、例えば、すべてのシーンについて同じパラダイムを用いて、同様に実行され得ることを理解され

40

50

たい。

【 0 0 9 2 】

次に、第 3 のドメインを決定するステップ S 2 0 の例を示すフローチャートを示す図 2 を参照して、第 3 のドメインを決定するステップ S 2 0 の例を説明する。

【 0 0 9 3 】

例では、第 3 のドメインを決定するステップ S 2 0 が仮想シーンのデータセットの各シーンについて、シーンから 1 つ又は複数の空間的に再構成可能なオブジェクトを抽出するステップ S 2 0 0 を含む。これらの例では、決定するステップ S 2 0 が抽出されたオブジェクトごとに、抽出されたオブジェクトを、データ分布に関して抽出されたオブジェクトよりも第 2 のドメインに近いオブジェクトに変換するステップ S 2 1 0 をさらに含む。

10

【 0 0 9 4 】

前述のように、仮想シーンのデータセットの仮想シーンは仮想シーンのデータセットを提供するステップ S 1 0 の前に、例えば、3 D シミュレーションソフトウェアを使用して、シーンが由来する仮想シミュレーションを計算することによって取得されてもよい。シミュレーションの計算は、各仮想シーンにおける各空間的に再構成可能なオブジェクトの自動位置特定及びアノテーション付けを含むことができる。データセットの各仮想シーンについて、及び仮想シーン内の各空間的に再構成可能なオブジェクトについて、ローカライゼーションは、空間的に再構成可能なオブジェクトの周りの境界ボックスの計算を含むことができる。境界ボックスは、シーンを表す画像の側に軸が平行な矩形ボックスである。境界ボックスは 4 つの座標によって特徴付けられ、空間的に再構成可能なオブジェクト全体を包含するように、適切な比率及びスケールで空間的に再構成可能なオブジェクトを中心とする。本開示のコンテキストにおいて、シーン内のオブジェクトを包含することはシーン上で完全に可視であるオブジェクトの全ての部分を包含することを意味することができ、前記部分はオブジェクト全体であるか、又はその一部のみである。アノテーションは各計算された境界ボックスに、境界ボックスが空間的に再構成可能なオブジェクトを包含することを示すラベルを割り当てることを含むことができる。ラベルは例えば、境界ボックスが産業用ロボットのような空間的に再構成可能な製造ツールを包含するなど、境界ボックスに関するさらなる情報を示すことができることを理解されたい。

20

【 0 0 9 5 】

1 つ又は複数の空間的に再構成可能なオブジェクトのそれぞれ 1 つについて、仮想シーンから空間的に再構成可能なオブジェクトを抽出するステップ S 2 0 0 は仮想シーンから空間的に再構成可能なオブジェクトを分離することになる任意のアクションを指定することができ、そのような分離を実行することができる任意の方法によって実行することができる。空間的に再構成可能なオブジェクトを仮想シーンから抽出するステップ S 2 0 0 は実際には空間的に再構成可能なオブジェクトを取り囲む境界ボックスを検出するステップを含むことができ、検出するステップは境界ボックスのラベルに基づいており、任意選択でラベルにアクセスするステップを含む。抽出するステップ S 2 0 0 はバウンディングボックスを仮想シーンから分離し、それによって、空間的に再構成可能なオブジェクトを仮想シーンから分離するステップをさらに含むことができる。

30

【 0 0 9 6 】

抽出されたオブジェクトを変換するステップ S 2 1 0 は、抽出されたオブジェクトを、データ分布に関して第 2 のドメインにより近い別のオブジェクトに変更する任意の変換を指定することができる。「第 2 のドメインに近い」ことは抽出されたオブジェクトへのオブジェクトがそれ自体に変換されることを意味し、それは第 1 のドメインよりも第 2 のドメインに近いドメインに属するシーンである。オブジェクトを抽出するステップ S 2 0 0 が前述のように、オブジェクトを取り囲む境界ボックスを仮想シーンから分離することを含む場合、抽出されたオブジェクトを変換するステップ S 2 1 0 は境界ボックス全体（したがって、抽出されたオブジェクトを含む）又は抽出されたオブジェクトのみの修正を含むことができ、境界ボックス全体又は抽出されたオブジェクトの修正は、データ分布に関して第 2 のドメインにそれを近づけるだけである。この変形例は境界ボックス又は空間的

40

50

に再構成可能なオブジェクトに、境界ボックスを用いて、仮想画像をデータ分布に関してより写実的な画像に変換することができるジェネレータを適用することを含むことができる。

【0097】

一実装形態では、変換するステップS210が空間的に再構成可能なオブジェクトの2つのセット、すなわち、仮想セット及び実セット上で訓練されたCycleGAN仮想対実ネットワーク([20]を参照)によって実行され得る。これは、仮想画像内の空間的に再構成可能なオブジェクト(例えば、ロボット)を修正するために使用される仮想から現実空間的に再構成可能なオブジェクト(例えば、ロボット)ジェネレータをもたらす。CycleGANモデルは、画像を1つのドメインから別のドメインに変換する。本発明は1つの画像セットの特別な特徴をキャプチャし、これらの特徴を第2の画像コレクションに変換する方法を、すべて対になった訓練例がない場合に、理解することができる方法を提案する。1対1マッピングを有するこの緩和は、実際にはこの定式化を非常に強力なものにする。これは生成モデル、具体的にはCycleGANと呼ばれる生成アドバーサルネットワーク(GAN)によって達成される。

【0098】

図19は、2つの仮想産業多関節ロボット190及び192を抽出するステップS200及び変換するステップS210の一例を示す。これらのロボット190及び192の各々は、それぞれの仮想シーンから抽出されている(S200)。抽出されたロボット190及び192は、次いで、それぞれ、よりフォトリアルなロボット194及び196に変換される(S210)。

【0099】

さらに図2のフローチャートを参照すると、第3のドメインの決定は変換された抽出オブジェクトごとに、変換された抽出オブジェクトを、データ分布に関して第1のドメインよりも第2のドメインにそれぞれ近い1つ又は複数のシーンに配置するステップS220をさらに含むことができる。そのような場合、第3のドメインは、変換された抽出されたオブジェクトが配置される各シーンを含む。

【0100】

1つ以上のシーンの各シーンは空間的に再構成可能なオブジェクトを含まないシーン、例えば、背景シーン(すなわち、背景自体が背景オブジェクトを含んでもよいが、単一の背景を含むシーン)であってもよい。1つ以上のシーンの各シーンはまた、変換された抽出されたオブジェクトが元々属するシーンと同じコンテキストからのシーンであってもよい。この同じコンテキストは、実際には仮想シーンのデータセットのすべてのシーンのコンテキストであってもよい。

【0101】

1つ又は複数のシーンの各シーンは、第1のドメインよりも第2のドメインに近い。「第2のドメインにより近い」とは、1つ以上のシーンの各シーンが第1のドメインよりも第2のドメインにより近いドメインに属することを意味する。例では、1つ又は複数のシーンの各シーンが前述のように現実から直接取得された現実のシーンである。代替的に、1つ以上のシーンの各シーンは仮想シーンのデータセットのそれぞれの仮想シーンから生成されてもよく、前記それぞれの仮想シーンの1つは、例えば、生成されたシーンがそれぞれの仮想シーンよりもリアル及び/又はフォトリアルスティックであるように、変換された抽出されたオブジェクトが属する仮想シーンである。データセットのそれぞれの仮想シーンからシーンを生成することは、仮想シーンに、仮想イメージを、現実及び/又はよりフォトリアルスティックなイメージに変換することができるジェネレータを適用することを含むことができる。一実施形態では、シーンの生成が2組のバックグラウンド(例えば、工場バックグラウンド)、すなわち、仮想セット及びリアルセットで訓練されたCycleGAN仮想・ツー・リアルジェネレータ([20]参照)によって実行され得る。これは、仮想画像内の背景(例えば、工場背景)を修正するために使用される仮想-実背景(例えば、工場背景)ジェネレータをもたらす。CycleGANモデル([20]参

10

20

30

40

50

照)は、画像を1つのドメインから別のドメインに変換する。本発明は1つの画像セットの特別な特徴をキャプチャし、これらの特徴を第2の画像コレクションに変換する方法を、すべて対になった訓練例がない場合に、理解することができる方法を提案する。1対1マッピングを有するこの緩和は、実際にはこの定式化を非常に強力なものにする。これは生成モデル、具体的にはCycleGANと呼ばれる生成アドバーサルネットワーク(GAN)によって達成される。

【0102】

変換された抽出されたオブジェクトを配置するステップS220は背景の画像にオブジェクトの単一及び/又は複数の画像を貼り付けることができる任意の方法によって(例えば、自動的に)実行されてもよい。したがって、変換された抽出オブジェクトの前記1つ又は複数のシーンを配置するステップS220は、そのような方法を適用して、変換された抽出オブジェクトを1つ又は複数のシーンのそれぞれ1つに貼り付けることを含むことができる。変換された抽出オブジェクトが置かれたシーンは、シーン上に既に置かれた1つ以上の他の変換された抽出オブジェクトを含んでもよいことが理解されるべきである(S220)。

【0103】

上述の例から、これらの例では、第3のドメインを決定するステップS20が特に、シーンの空間的に再構成可能なオブジェクトと、それらが属するシーンの背景とを別々に変換し、次に、変換された空間的に再構成可能なオブジェクトのそれぞれを、変換された背景のうちの1つ又は複数の背景に配置することができることになる。第1に、これは、仮想シーンのデータセットの仮想シーンを完全に(すなわち、空間的に再構成可能なオブジェクトとそれらの背景とを分離することなく)変換された背景において変換されたオブジェクトのより多くの構成を得ることを可能にする。ドメイン適応ニューラルネットワークを学習するステップS30は、第3のドメインのシーンに基づくので、学習はよりロバストである。第2に、仮想シーンのデータセットの仮想シーン全体を変換することは、実背景の変動性のためにテストドメインのシーン間に基礎をなす関係がない可能性があるため、不十分な結果につながる可能性がある。例えば、空間的に再構成可能なオブジェクトを含む仮想シーン全体を変換することは、空間的に再構成可能なオブジェクトが背景から区別され得ない(例えば、背景と混合され得る)変換されたシーンをもたらす得る。これらの全ての理由のために、仮想シーンのデータセットの仮想シーンを完全に変換することは、例えば、第1のドメインよりも第2のドメインに近づくことができない第3のドメインを生成することによって、不十分な結果をもたらす可能性がある。さらに、この場合、第3のドメインのシーンは、これらのシーンにおける関心のあるオブジェクト(空間的に再構成可能なオブジェクト)の特性が変換によって影響を受ける可能性があるため、学習に適していない可能性がある。

【0104】

図20は、変換された抽出オブジェクトを配置するステップS220の一例を示す図である。図20は、工場背景技術シーン200及び202にそれぞれS220を配置した図19の例の変換された抽出された産業用関節ロボット194及び196を示す。

【0105】

さらに図2のフローチャートを参照すると、例では、変換され抽出されたオブジェクトの1つ又は複数のシーンの配置するステップS220がランダムに実行される。変換された抽出オブジェクトのそのようなランダム配置ステップS220は背景の画像内のオブジェクトの単一及び/又は複数の画像をランダムに貼り付けることができる任意の方法によって(例えば、自動的に)実行されてもよい。したがって、変換された抽出オブジェクトをランダム配置するステップS220は、そのような方法を適用して、変換された抽出オブジェクトを1つ又は複数のシーンのそれぞれ1つにランダムに貼り付けることを含むことができる。

【0106】

変換され抽出されたオブジェクトをランダムに配置するステップS220は、第3のド

10

20

30

40

50

メインのシーンからなる学習セットに基づくドメイン適応ニューラルネットワークを学習するステップS 3 0のロバスト性を改善する。実際に、変換された抽出オブジェクトをランダムに配置することは学習するステップS 3 0の間に変換された抽出オブジェクトの位置にドメイン適応型ニューラルネットワークが焦点を当てないことを保証するが、それは学習するステップS 3 0がそれらに頼るため、及び/又は変換された抽出オブジェクトとそれらの位置との間の相関を学習するために、そのような位置のあまりにも広いスパン範囲が存在するからである。代わりに、学習するステップS 3 0は空間的に再構成可能なオブジェクトの推論に、それらの位置にかかわらず、真に焦点を当てることができる。なお、空間的に再構成可能なオブジェクトの位置に依存しないことは、単にオブジェクトが適切な位置を占有しないという理由で、ドメイン適応型ニューラルネットワークが推論すると想定される空間的に再構成可能なオブジェクトを、ドメイン適応型ニューラルネットワークが見逃すリスクを大幅に低減することを可能にする。

10

【0107】

上記の例では、第3のドメインが変換された抽出オブジェクトごとに、変換された抽出オブジェクトが配置(S 2 2 0)される各シーンを含む。例えば、第3のドメインは、このようなシーンで構成される。すなわち、仮想シーンのデータセットの全てのシーンについて、抽出するステップS 2 0 0、変換するステップS 2 1 0、及び配置するステップS 2 2 0が同様に行われる。例えば、これらは、1つ以上の変換された抽出されたオブジェクトが配置(S 2 2 0)される全てのシーンが実質的に同じレベルのフォトリアリズムを有するように実行される。言い換えると、それらはドメイン、すなわち第3のドメインに属し、及び/又は形成する。

20

【0108】

さらに図2のフローチャートを参照すると、第3のドメインを決定するステップS 2 0は、第3のドメインに含まれる1つ以上のシーンのそれぞれ1つに1つ以上のディストラクタを配置するステップS 2 3 0を含むことができる。ディストラクタは、空間的に再構成可能なオブジェクトではないオブジェクトである。

【0109】

ディストラクタは、変換された抽出されたオブジェクトと実質的に同じレベルのフォトリアリズムを有するオブジェクトであってもよい。実際には、ディストラクタが仮想シーンの空間的に再構成可能なオブジェクトと同様に、前記仮想シーンから抽出され、第2のドメインにより近いオブジェクトに変換され、第3のドメインのシーンに配置される仮想シーンのデータセットの仮想シーンのオブジェクトであってもよい。あるいはディストラクタが現実の又は仮想のディストラクタの別のデータセットから生じてもよく、この場合、決定するステップS 2 0はそのようなデータセットを提供するステップを含んでもよい。いずれにせよ、ディストラクタは、第3のドメインのシーンに配置(S 2 3 0)されてもよく、そのような配置ステップS 2 3 0の後、シーンが依然として第3のドメインに属するようなフォトリアリズムのレベルを有することを理解されたい。

30

【0110】

ディストラクタは例えば、前述のように仮想製造シーンから抽出され変換された機械的部品、又は、例えば、仮想(例えば、製造)シーンの背景の一部である任意の他のオブジェクトである。図21は、シーンを製造する状況において第3のドメインの1つ又は複数のシーンに配置される3つのディストラクタ210、212、及び214を示す。ディストラクタ210、212、及び214はそれぞれ、ハードハット(製造シーンの背景の対象物である)、金属シート(機械部品である)、及び鉄馬蹄形の鍛造部品(機械部品である)である。

40

【0111】

ディストラクタ、特に変換された抽出オブジェクトと同様に得られるディストラクタを(前述のように)配置するステップS 2 3 0により、数値(例えば、デジタル、例えば、画像化)アーチファクトに対するニューラルネットワークの感度を低減することができる。実際に、抽出するステップS 2 0 0及び/又は変換するステップS 2 1 0及び/又は配

50

置するステップS 2 2 0の結果として、少なくとも1つの変換された抽出オブジェクトを特徴第3のドメインのシーンは、少なくとも1つの変換された抽出オブジェクトが配置(S 2 2 0)される場所に1つ又は複数の数値アーティファクトを特徴付けることができる。シーン内にディストラクタを配置するステップS 2 3 0は、ディストラクタが配置された場所に同じ種類のアーチファクトを出現させ得る。したがって、学習(S 3 0)されると、ドメイン適応ニューラルネットワークは、アーチファクトに焦点を当てず、又は少なくとも、アーチファクトがディストラクタによっても引き起こされるので、アーチファクトを空間的に再構成可能なオブジェクトにリンクしない。これは、学習するステップS 3 0をよりロバストにする。

【0 1 1 2】

図1のフローチャートに戻って、ドメイン適応ニューラルネットワークを学習するステップS 3 0を説明する。

【0 1 1 3】

「機械学習」と「ニューラルネットワークの学習」の一般的な概念を最初に簡単に論じた。

【0 1 1 4】

機械学習の分野から知られているように、ニューラルネットワークによる入力処理は入力に動作を適用することを含み、動作は、重み値を含むデータによって定義される。したがって、ニューラルネットワークの学習はそのような学習のために構成されたデータセットに基づいて重みの値を決定することを含み、このようなデータセットは、学習データセット又は訓練データセットと呼ばれる可能性がある。そのために、データセットは、各々がそれぞれの訓練サンプルを形成するデータピースを含む。訓練サンプルは、ニューラルネットワークが学習された後に使用される状況の多様性を表す。本明細書で言及される任意のデータセットは、1 0 0 0、1 0 0 0 0、1 0 0 0 0 0、又は1 0 0 0 0 0 0を超える訓練サンプルの数を含むことができる。本開示のコンテキストでは「データセットに基づいてニューラルネットワークを学習する」とはデータセットがニューラルネットワークの学習/訓練セットであることを意味し、「ドメインに基づいてニューラルネットワークを学習する」とはニューラルネットワークの訓練セットがドメインに属し、及び/又はドメインを形成するデータセットであることを意味する。

【0 1 1 5】

本開示の任意のニューラルネットワーク、例えば、ドメイン適応ニューラルネットワーク、又は検出器、教師エクストラクタ、又は生徒エクストラクタはすべて後述されるが、任意の深層学習技術によって学習される深層ニューラルネットワーク(DNN)であってもよい。深層学習技術はニューラルネットワークにおける学習のための強力な一連の技術であり([19]参照)、これはコンピュータが観察データから学習することを可能にする生物学的にインスパイアされたプログラミングパラダイム(biologically-inspired programming paradigm)である。画像認識において、DNNの成功は他の画像分類法(SVM、Boosting、Random Forest)で使用される手で設計された低レベル特徴(Zernikeモーメント、HOG、Bag of Words、SIFT)とは対照的に、豊富な中レベルメディア表現を学習する能力に起因する。より具体的には、DNNが生データに基づくエンドツーエンド学習に焦点を当てている。言い換えれば、それらは、生の特徴で始まりラベルで終わるエンドツーエンド最適化を達成することによって、可能な最大限まで特徴エンジニアリングから離れる。

【0 1 1 6】

ここで、「シーン内の空間的に再構成可能なオブジェクトの推論のために構成される」という概念について説明する。この議論は特に、ドメイン適応型ニューラルネットワークが学習される方法とは独立して、ドメイン適応型学習方法、ドメイン適応型推論方法、又はプロセスのいずれか1つによるドメイン適応型ニューラルネットワークに適用される。シーン内の空間的に再構成可能なオブジェクトを推論することは、そのようなオブジェクトのうちの1つ又は複数を含むシーン内の1つ又は複数の空間的に再構成可能なオブジェ

10

20

30

40

50

クトを推論することを手段する。これは、シーン内の全ての空間的に再構成可能なオブジェクトを推論することを意味することができる。したがって、シーン内の空間的に再構成可能なオブジェクトを推論するように構成されたニューラルネットワークはシーンを入力として受け取り、空間的に再構成可能なオブジェクトのうちの1つ又は複数の位置に関するデータを出力する。例えば、ニューラルネットワークはシーンの（例えば、デジタル）画像表現を出力することができ、それぞれの境界ボックスは、推論された空間的に再構成可能なオブジェクトのそれぞれ1つをそれぞれ包含する。推論はまた、そのような境界ボックスの計算、及び／又は境界ボックスのラベル付け（例えば、シーンの空間的に再構成可能なオブジェクトを包含するものとして）、及び／又は推論された空間的に再構成可能なオブジェクトのラベル付け（空間的に再構成可能なオブジェクトとして）を含むことができる。ニューラルネットワークの出力は、データベースに記憶され、及び／又はコンピュータのスクリーン上に表示されてもよい。

10

【0117】

ドメイン適応ニューラルネットワークを学習するステップS30は、現実のシーンにおける空間的に再構成可能なオブジェクトの推論のために構成されたニューラルネットワークを学習することができる任意の機械学習アルゴリズムによって実行されてもよい。

【0118】

ドメイン適応ニューラルネットワークを学習するステップS30は、第3のドメインに基づく。したがって、第3のドメインは、ドメイン適応ニューラルネットワークの訓練／学習セットであるか、又は少なくともそれを含む。第3のドメインを決定するステップS20に関する前の議論から、ドメイン適応ニューラルネットワークシステムの訓練／学習セット（例えば、得られる）が、仮想シーンの提供されたS10データセットから得られることが明らかである。言い換えれば、決定するステップS20は仮想シーンのデータセットである理論的訓練／学習セットを、一旦学習されたニューラルネットワークがその入力を取得するセットにより近い、実用的な訓練／学習セット、すなわち第3のドメインのシーンに変換するための、理論的訓練／学習セットの前処理と考えることができる。これは、上述したように、学習するステップS30のロバスト性及びドメイン適応ニューラルネットワークの出力品質を改善する。

20

【0119】

ドメイン適応型ニューラルネットワークはドメイン適応型ニューラルネットワークの少なくとも一部を形成する、オブジェクト検出器とも呼ばれる検出器を備えてもよい（例えば、それから構成されてもよい）。したがって、検出器はそれ自体ニューラルネットワークであり、さらに、現実のシーン内の空間的に再構成可能なオブジェクトを検出するように構成される。検出器は、典型的には何百万ものパラメータ及び／又は重みを含むことができる。ドメイン適応方法を学習するステップS30は、これらのパラメータ及び／又は重みの値を設定する検出器の訓練を含み得る。そのような訓練は典型的にはこれらの値を更新することを含み得、それは検出器によって取られた各入力に対する検出器の出力に従って、これらの値を連続的に補正することから成る。訓練中、検出器の入力はシーンであり、出力はこのシーン内の1つ又は複数の（例えば、すべての）空間的に再構成可能なオブジェクト（例えば、前述のように、境界ボックス及び／又はアノテーション）の位置に関する（少なくとも適切な訓練の後の）データを含む。

30

40

【0120】

値の訂正は、各入力に関連するアノテーションに基づいてもよい。アノテーションは、モデルの出力が真又は偽の場合に評価を可能にする、特定の入力に関連付けられたデータのセットである。例えば、前述のように、第3のドメインに属するシーンの各空間的に再構成可能なオブジェクトは、境界ボックスによって包含されてもよい。前述のように、本開示のコンテキストでは、シーン内のオブジェクトを包含することはシーン上で完全に可視であるオブジェクトのすべての部分を包含することを意味することができる。境界ボックスは、それが空間的に再構成可能なオブジェクトを包含することを示すようにアノテーション付けされてもよい。したがって、値の補正は、検出器に入力された第3のドメイン

50

シーンの境界ボックスのそのようなアノテーションを評価することと、評価されたアノテーションに基づいて、境界ボックスに含まれる空間的に再構成可能なオブジェクトが実際に検出器によって出力されるかどうかを判定することとを含むことができる。検出器の出力が境界ボックス内に含まれる空間的に再構成可能なオブジェクトを含むかどうかを判定することは、例えば、出力が真であるかどうか（例えば、それが空間的に再構成可能なオブジェクトを含む場合）、又は偽であるかどうか（例えば、それが空間的に再構成可能なオブジェクトを含まない場合）を評価することによって、検出器の出力と空間的に再構成可能なオブジェクトとの間の対応関係を評価することを含むことができる。

【0121】

仮想データセットのアノテーションを使用して検出器を訓練する上記の方法は「監視学習」と呼ばれることがある。検出器の訓練が任意の監視学習方法によって実行されることがある。検出器が訓練されると、値の補正は停止する。この時点で、検出器は新たな入力（すなわち、検出器の訓練中の未知の入力）を処理し、検出結果を返すことができる。本発明のコンテキストでは検出器の新たな入力テストドメインのシーンである。訓練されると、検出器は認識（又は分類）のタスクと、テストデータセットの実シーンにおける空間的に再構成可能なオブジェクトの位置特定のタスクとを共同で実行する「検出」手段のタスクとして、2つの異なった出力を返す。

【0122】

位置特定のタスクは、検出器に入力された現実のシーンの空間的に再構成可能なオブジェクトを各々が包含する境界ボックスを計算することを含む。前述のように、境界ボックスは軸が画像の側に平行である矩形ボックスであり、4つの座標によって特徴付けられる。例では、検出器が検出器に入力されたシーンの各空間的に再構成可能なオブジェクトについて、適切な比率及びスケールでオブジェクトを中心とする境界ボックスを返す。

【0123】

分類のタスクは計算された各境界ボックスに対応するラベル（又はアノテーション）をラベル付けし、ラベルに信頼スコアを関連付けることを含む。信頼性スコアは、境界ボックスが入力として提供されるシーンの空間的に再構成可能なオブジェクトを真に包含するという検出器の信頼性を反映する。信頼度スコアは、0と1との間の実数であってもよい。このような場合、信頼スコアが1に近ければ近いほど、検出器は、対応する境界ボックスに関連付けられたラベルが空間的に再構成可能なオブジェクトに真にアノテーションを付けることについての信頼度が高くなる。

【0124】

次に、ドメイン適応ニューラルネットワークを学習するステップS30の例を示すフローチャートを示す図3を参照して、ドメイン適応ニューラルネットワークを学習するステップS30の例を説明する。

【0125】

図3のフローチャートを参照すると、ドメイン適応ニューラルネットワークを学習するステップS30は、教師エクストラクタを提供するステップS300を含むことができる。教師エクストラクタは、実シーンの画像表現を出力するように構成された機械学習ニューラルネットワークである。そのような場合、ドメイン適応ニューラルネットワークを学習するステップS30は、生徒エクストラクタを訓練することS310をさらに含む。生徒エクストラクタは、第3のドメインに属するシーンの画像表現を出力するように構成されたニューラルネットワークである。生徒エクストラクタを訓練するステップS310は、損失を最小化することを含む。1つ以上の現実シーンの各シーンについて、損失は、教師のエクストラクタをシーンに適用した結果と、生徒のエクストラクタをシーンに適用した結果との間の格差をペナルティにする。

【0126】

生徒エクストラクタは、第3のドメインに属するシーンの画像表現を出力するように構成されたニューラルネットワークである。生徒エクストラクタは検出器の一部であってもよく、したがって、生徒エクストラクタは第3のドメインに属するシーンの前記画像表現

10

20

30

40

50

を検出器の他の部分に出力するようにさらに構成されてもよい。そのような他の部分のうちの少なくとも1つはローカライゼーションのタスクのために構成されてもよく、そのような他の部分のうちの少なくとも1つは前述のように、分類のタスクのために構成されてもよい。したがって、生徒エクストラクタは第3のドメインのシーンの画像表現を前記他の部分に出力し、前記他の部分がローカライゼーション及び分類のタスクを実行することができるようにすることができる。あるいは、生徒エクストラクタがドメイン適応ニューラルネットワークの一部であってもよく、この場合、画像表現をドメイン適応ニューラルネットワークの他の部分（検出器など）に出力してもよい。

【0127】

これは、生徒エクストラクタが第3のドメインに属するシーンの画像表現の出力を監視し、一方、ドメイン適応ニューラルネットワークの少なくとも1つの他の部分（例えば、前述の検出器）が第3のドメインの各シーンにおける空間的に再構成可能なオブジェクトの検出（例えば、局所化及び分類）を監視することを意味する。前記検出は、生徒エクストラクタによって出力された前記画像表現に基づくことができる。換言すれば、生徒エクストラクタによって出力される画像表現は、ドメイン適応ニューラルネットワークの残りの部分、例えば生徒エクストラクタが検出器の一部を形成する場合の検出器の残りの部分に供給される。

【0128】

生徒エクストラクタは、提供（S300）された教師エクストラクタの助けを借りて訓練される。教師エクストラクタは、実シーンの画像表現を出力するように構成された機械学習ニューラルネットワークである。実施例では、教師エクストラクタが例えば、実データスタイルのためのロバストな畳み込みフィルタを学習することによって、実シーン上で学習されて、実シーンの画像表現を出力するという手段がある。教師エクストラクタのニューロンの層は、典型的には出力されるべき実画像の表現を符号化する。教師エクストラクタは、例えばオープンソースから利用可能な現実シーンの任意のデータセット上で、及び任意の機械学習技術によって学習されていてもよい。教師エクストラクタを提供するステップS300は、教師エクストラクタが学習した後に格納されているデータベースにアクセスするステップと、データベースから教師エクストラクタを検索するステップとを含むことができる。教師エクストラクタは既に利用可能であってもよく、すなわち、ドメイン適応学習方法は、既に利用可能な教師エクストラクタを提供するステップS300のみによる教師エクストラクタの学習を含まなくてもよいことを理解されたい。あるいは、ドメイン適応学習方法が教師エクストラクタを提供するステップS300の前に、任意の機械学習技術による教師エクストラクタの学習を含むことができる。

【0129】

さらに図2のフローチャートを参照して、生徒エクストラクタを訓練するステップS320について説明する。

【0130】

前記訓練ステップS320は、1つ又は複数の現実のシーンの各シーンについて、シーンに教師エクストラクタを適用した結果とシーンに生徒エクストラクタを適用した結果との間の不一致にペナルティを課す損失を最小限に抑えることを含む。1つ以上の現実シーンの少なくとも一部（例えば、全て）は、テストデータセットのシーンであってもよく、及び/又は1つ以上の現実シーンの少なくとも一部（例えば、全て）は第2のドメインに属する現実シーンの別のデータセットのシーンであってもよく、例えば、アノテーションが付けられておらず、必ずしも多数ではない。損失の最小化は教師エクストラクタと生徒エクストラクタの両方を、それぞれの1つ以上の現実シーンに適用した結果を含み、それは、手段、実シーンが後者を訓練するステップS310の間に教師エクストラクタと生徒エクストラクタの両方に与えられることを意味する。したがって、ドメイン適応型ニューラルネットワークの少なくとも一部は前述の検出器などの第3のドメインに属する仮想データ上で訓練されるが、ドメイン適応型ニューラルネットワークの少なくとも一部、すなわち生徒エクストラクタは実データ上で訓練される。

10

20

30

40

50

【 0 1 3 1 】

損失は1つ又は複数の現実のシーンの各現実のシーンについて、シーンに教師エクストラクタを適用した結果とシーンに生徒エクストラクタを適用した結果との間の類似性及び／又は非類似性を測定する量（例えば、関数）とすることができる。損失は例えば、1つ以上の現実シーンの各シーンをパラメータとしてとる関数であってもよく、入力として、シーンに教師エクストラクタを適用した結果と、シーンに生徒エクストラクタを適用した結果とを取り込み、シーンに教師エクストラクタを適用した結果とシーンに生徒エクストラクタを適用した結果との間の類似性及び／又は非類似性を表す量（陽性の実数など）を出力する。1つ以上の現実シーンの各シーンについて、教師エクストラクタをシーンに適用した結果と生徒エクストラクタをシーンに適用した結果との間の差異は、教師エクストラクタをシーンに適用した結果と生徒エクストラクタをシーンに適用した結果との間の差異の定量化である。不均衡をペナライズすることは、損失が不均衡の増加関数であることを意味し得る。

10

【 0 1 3 2 】

教師エクストラクタと生徒エクストラクタの出力間の不一致にペナルティを課すことによって損失を最小化することは、ドメイン適応ニューラルネットワークを学習するステップS30のロバスト性を改善する。特に、生徒エクストラクタを訓練するステップS320中に、既に学習されており、したがって、重み及び／又はパラメータが変更されない教師エクストラクタは、生徒エクストラクタをガイドする。実際、生徒エクストラクタの出力と教師エクストラクタの出力との間の不一致はペナルティを課されるので、生徒エクストラクタがデータ分布に関して教師エクストラクタによって出力されたものに近い実画像表現を出力することを学習するように、生徒エクストラクタのパラメータ及び／又は重みが修正される。言い換えれば、生徒のエクストラクタは、実画像表現を出力する方法で教師のエクストラクタを模倣するように訓練されている。したがって、生徒エクストラクタは、教師エクストラクタの訓練中に教師エクストラクタによって学習されたものと同じ又は実質的に同じ実データスタイルのロバストな畳み込みフィルタを学習することができる。その結果、生徒エクストラクタは、これらのシーンがデータ分布に関して第2のドメイン（現実シーンからなる）から比較的離れていても、元々第3のドメインに属するシーンの現実画像表現を出力する。これは、生徒エクストラクタが訓練（S320）される方法が第3のドメインのアノテーション付き仮想シーン上で訓練されている間に、ドメイン適応ニューラルネットワークが実シーンから空間的に再構成可能なオブジェクトを推論することを可能にするので、ドメイン適応ニューラルネットワーク、又は適切な場合にはその検出器をデータ外観に対して不変にすることを可能にする。

20

30

【 0 1 3 3 】

損失は「蒸留損失」と呼ばれてもよく、教師エクストラクタを使用する生徒エクストラクタを訓練するステップS320は「蒸留アプローチ」と呼ばれてもよい。蒸留損失の最小化は、任意の最小化アルゴリズム又は任意の緩和アルゴリズムによって実施することができる。蒸留損失は、ドメイン適応ニューラルネットワークの損失の一部であってもよく、任意選択で、前述の検出器の損失も含む。一実施形態では、ドメイン適応型ニューラルネットワークを学習するステップS30が以下のタイプのドメイン適応型ニューラルネットワークの損失を最小限に抑えることを含むことができる。

40

【 数 1 7 】

$$L_{training} = L_{detector} + L_{distillation}$$

ここで、

【 数 1 8 】

$$L_{training}$$

はドメイン適応ニューラルネットワークを学習するステップ S 3 0 に対応する損失であり、
【数 1 9】

$$L_{detector}$$

は検出器の訓練に対応する損失であり、
【数 2 0】

$$L_{distillation}$$

は蒸留損失である。したがって、ドメイン適応ニューラルネットワークの損失
【数 2 1】

$$L_{training}$$

を最小化することは、任意の緩和アルゴリズムによって、検出器の損失
【数 2 2】

$$L_{detector}$$

及び蒸留損失
【数 2 3】

$$L_{distillation}$$

を同時に又は独立して最小化することを含んでもよい。
【0 1 3 4】

例では、シーンに教師エクストラクタを適用した結果が第 1 のグラム行列であり、シーンに生徒エクストラクタを適用した結果が第 2 のグラム行列である。エクストラクタ（例えば、生徒又は教師）のグラム行列は、エクストラクタによって出力される画像上のデータの分布の表現形式である。より具体的には、エクストラクタをシーンに適用することにより画像表現が生成され、画像表現のグラム行列を計算することにより、画像表現の分布の埋め込み、いわゆる画像のグラムベース表現が生成される。これは例では教師エクストラクタをシーンに適用した結果が教師エクストラクタによって出力された画像表現の第 1 のグラム行列の計算から得られる教師エクストラクタによって出力された画像表現の第 1 のグラムベースの表現であり、生徒エクストラクタをシーンに適用した結果が生徒エクストラクタによって出力された画像表現の第 2 のグラム行列の計算から得られる生徒エクストラクタによって出力された画像表現の第 2 のグラムベースの表現でということになる。例えば、グラム行列はテクスチャ、形状及び重みのような、エクストラクタによって出力された画像についての非ローカライズ情報を含み得る。

【0 1 3 5】

例では、第 1 のグラム行列が教師エクストラクタのニューロンのいくつかの層上で計算される。ニューロンの前記層は、少なくともニューロンの最後の層を含む。これらの例では、第 2 のグラム行列が生徒エクストラクタのニューロンのいくつかの層上で計算される。ニューロンの前記層は、少なくともニューロンの最後の層を含む。

【 0 1 3 6 】

エクストラクタ（教師又は生徒）のニューロンの最後の層は、エクストラクタによって出力された画像を符号化する。1 番目と 2 番目のグラム行列はともに、ニューロンのより深層層である数層のニューロン上で計算される。この手段は、両方とも、画像自身よりも出力された画像について多くの情報を含むことである。これにより、生徒エクストラクタの訓練がより頑健になる。なぜなら、それは教師と生徒エクストラクタの出力間のリアルスタイル近接度のレベルを高めるからであり、これは損失の最小化が目指すものである。特に、生徒エクストラクタはそれによって、その教師と同様に、実画像の 2 つのレイヤ表現を正確に推論することができる。

10

【 0 1 3 7 】

例では、視差がシーンに教師エクストラクタを適用した結果と、シーンに生徒エクストラクタを適用した結果との間のユークリッド距離である。

【 0 1 3 8 】

ここで、蒸留損失の最小化の実施について説明する。この実施形態では、ニューロン

【数 2 4 】

ℓ

20

のあるレイヤにおいてエクストラクタ（教師又は生徒）によって出力される画像のスタイル表現がこのレイヤにおける画像の特徴の分布によって符号化される。この実施形態では、特徴分布埋め込みの一形態である特徴相関を符号化するためにグラム行列が使用される。形式的には長さ

【数 2 5 】

N^ℓ

30

のフィルタを有する応答

【数 2 6 】

F^ℓ

の層

【数 2 7 】

ℓ

40

について、特徴相関はグラム行列

【数 2 8 】

$G^\ell \in R^{N^\ell \times N^\ell}$

によって与えられ、ここで、

【数 2 9 】

$G_{i,j}^\ell$

50

は層

【数 3 0】

ℓ

内のベクトル化された特徴マップ i と j との間の内積である。

【数 3 1】

$$G_{i,j}^{\ell} = \sum_k F_{i,k}^{\ell} F_{j,k}^{\ell}$$

10

【0 1 3 9】

提案された蒸留アプローチの背後にある直感は学習中に、生徒エクストラクタの重みが、教師エクストラクタが実データに対して生成するものに近いデータスタイル表現を生徒エクストラクタが出力するように更新されることを保証することである。そうするために、学習するステップ S 3 0 は、特徴エクストラクタレベルでグラムベースの蒸留損失

【数 3 2】

$L_{distillation}$

20

を最小化することを含む。実画像が与えられ、

【数 3 3】

F^{ℓ} 及び E^{ℓ}

30

を、生徒エクストラクタ及び教師エクストラクタのそれぞれのある層

【数 3 4】

ℓ

における対応する特徴マップ出力とする。したがって、層

【数 3 5】

ℓ

40

において提案された蒸留損失は、次のタイプのものである。

【数 3 6】

$$L_{distillation}^{\ell} = \|G_F^{\ell} - G_E^{\ell}\|,$$

ここで、

50

【数 3 7】

$$\|G_F^\ell - G_E^\ell\|$$

は対応するグラム行列

【数 3 8】

$$G_F^\ell \text{ と } G_E^\ell$$

10

との間の任意の距離（例えば、ユークリッド距離）である。層

【数 3 9】

$$\{\ell_1, \dots, \ell_m\}$$

知識を蒸留することを考慮すると、蒸留損失は次のタイプのものである。

【数 4 0】

$$L_{distillation} = \sum_{1 \leq i \leq m} L_{distillation}^{\ell_i}$$

20

【0 1 4 0】

前述のように、ドメイン適応ニューラルネットワークを学習するステップ S 3 0 は、以下のタイプの損失を最小化することを含むことができる：

【数 4 1】

$$L_{training} = L_{detector} + L_{distillation}$$

30

【0 1 4 1】

図 1 のフローチャートに戻って、中間ドメインを決定するステップ S 5 0 について説明する。

【0 1 4 2】

中間ドメインを決定するステップ S 5 0 はデータ分布の観点から、中間ドメインがテストドメインよりも訓練ドメインに近いようなものである。したがって、中間ドメインを決定するステップ S 5 0 はシーンのテストデータセットの各シーンについて、別のシーンの推論を指定し、その結果、推論された別のシーンのすべてが、データ分布に関してテストドメインよりも訓練ドメインに近いドメイン、すなわち中間ドメインを形成する。前記別のシーンの推論は前記別のシーンへの前記シーンの変換（例えば、変換）を含むことができる。変換は例えば、画像を 1 つのドメインから別のドメインに転送することができるジェネレータを使用して、シーンを前記別のシーンに変換することからなってもよい。そのような変換はシーンの 1 つ又は複数の部分のみに関係し、前記 1 つ又は複数の部分は例えば、ジェネレータを使用して変換され、一方、シーンの残りの部分は、変更されないままであることを理解されたい。前記別のシーンは、1 つ以上の部分が変換され、残りのシーンは変更されないシーンから構成されてもよい。あるいは、変換がシーンの残りを（例えば、訓練ドメインに属する）他のシーンの一部で置き換えることを含んでもよい。そのような場合、前記別のシーンは、1 つ又は複数の部分が変換され、残りのシーンが何らかの他のシーンの前記部分によって置き換えられているシーン内にあってもよい。推論は例え

40

50

ば、全てのシーンに対して同じパラダイムを用いて、テストドメインの全てのシーンに対して同様に実行されてもよいことが理解されるべきである。

【 0 1 4 3 】

ドメイン適応ニューラルネットワークは、訓練ドメインではなく、訓練ドメインから得られたデータで学習される。このようなデータは、訓練ドメインのデータの処理の結果であってもよい。例えば、訓練ドメインから得られるデータは、訓練ドメインのそれぞれのシーンを変換することによってそれぞれ得られるシーンのデータセットであってもよい。これは、訓練ドメインから得られたデータがドメイン適応ニューラルネットワークの真の訓練 / 学習ドメインであるドメインに属するシーンのデータセットを形成し得ることを意味する。このようなドメインは第 3 のドメインとして正確に決定されてもよく、すなわち、同様の動作及びステップを実行することによって決定されてもよい。いずれの場合も、中間ドメインを決定するステップ S 5 0 は、中間ドメインがテストドメインよりも、訓練ドメインから取得されたデータによって形成されたドメインに近いように（例えば、前述のようなシーンの変換を通じて）実行されてもよい。例では、訓練ドメインから取得されたデータがデータ分布に関して訓練ドメインよりも中間ドメインに近い別の中間ドメインを形成する。前記別の中間ドメインは、前述の第 3 のドメインであってもよい。これらのすべての例において、中間ドメインを決定するステップ S 5 0 はドメイン適応ニューラルネットワークが学習された真の訓練 / 学習セットに近づくように、テストデータセットの前処理に相当する。前述したように、これは、ドメイン適応ニューラルネットワークの出力品質を改善する。

【 0 1 4 4 】

例では、中間ドメインを決定するステップ S 5 0 が試験データセットの各シーンについて、試験データセットのシーンを、データ分布に関して訓練ドメインにより近い別のシーンに変換することを含む。

【 0 1 4 5 】

シーンを別のシーンに変換することは、画像を 1 つのドメインから別のドメインに転送することができるジェネレータを使用することを含むことができる。そのような変換はシーンの 1 つ又は複数の部分のみに関係し、前記 1 つ又は複数の部分は例えば、ジェネレータを使用して変換され、一方、シーンの残りの部分は、変更されないままであることを理解されたい。前記別のシーンは、1 つ以上の部分が変換され、残りのシーンは変更されないシーンから構成されてもよい。あるいは、変換がシーンの残りを（例えば、訓練ドメインに属する）他のシーンの一部で置き換えることを含んでもよい。そのような場合、前記別のシーンは、1 つ又は複数の部分が変換され、残りのシーンが何らかの他のシーンの前記部分によって置き換えられているシーン内であってもよい。推論は例えば、全てのシーンに対して同じパラダイムを用いて、テストドメインの全てのシーンに対して同様に実行されてもよいことが理解されるべきである。

【 0 1 4 6 】

次に、中間ドメインを決定するステップ S 5 0 の例を示すフローチャートを示す図 4 を参照して、中間ドメインを決定するステップ S 5 0 の例を説明する。

【 0 1 4 7 】

図 4 のフローチャートを参照すると、シーンの変換は、テストデータセットのシーンから仮想シーンを生成するステップ S 5 0 0 を含むことができる。このような場合、別のシーンは、テストデータセットのシーン及び生成された仮想シーンに基づいて推論される。

【 0 1 4 8 】

生成するステップ S 5 0 0 は、テストドメインのシーンから仮想シーンを生成することができる任意のジェネレータを適用することによって実行することができる。別のシーンはテストデータセットのシーンと生成された仮想シーンとを混合することによって、及び / 又はテストデータセットのシーンと生成された仮想シーンとを結合する（例えば、混合する、シームレスクローニングする、又は、例えば、マージする画像チャネル方法によってマージする）ことによって推論されてもよい。

【 0 1 4 9 】

さらに図4のフローチャートを参照すると、テストデータセットのシーンから仮想シーンを生成するステップS500は、テストデータセットのシーンに仮想シーンジェネレータを適用するステップを含むことができる。そのような場合、仮想シーンジェネレータは、テストドメインのシーンから仮想シーンを推論するように構成された機械学習ニューラルネットワークである。

【 0 1 5 0 】

仮想シーンジェネレータは入力として、テストドメインのシーンを取り込み、仮想シーン、例えば、テストドメインのシーンよりもフォトリアリスティックでない仮想シーンを出力する。仮想ジェネレータは、既に訓練/学習されていてもよいことを理解されたい。したがって、仮想シーンを生成するステップS500は例えば、仮想シーンジェネレータの訓練/学習後に仮想シーンジェネレータが記憶されているメモリにアクセスし、メモリから仮想シーンジェネレータを取り出すことによって、仮想シーンジェネレータを提供することを含むことができる。加えて、又は代替として、仮想シーンジェネレータの提供は、仮想シーンジェネレータの学習/訓練によって、そのような訓練に適した任意の機械学習技術によって先行されてもよい。

【 0 1 5 1 】

例では、仮想シーンジェネレータが各々1つ又は複数の空間的に再構成可能なオブジェクトを含むシーンのデータセット上で学習済である。そのような例では、仮想シーンジェネレータが1つ又は複数の空間的に再構成可能なオブジェクトを含むシーンが供給されたときに、1つ又は複数の空間的に再構成可能なオブジェクト、例えばすべてが同じ位置及び/又は構成にあるオブジェクトも含む仮想シーンであるシーンの仮想化を出力することができる。前述のように、ドメイン適応推論方法及び/又はプロセスは仮想シーンジェネレータを提供することのみを含むことができ(すなわち、その学習なしに)、又は仮想シーンジェネレータの学習をさらに含むことができることを理解されたい。

【 0 1 5 2 】

ドメイン適応推論方法及び/又は処理の実施において、テストデータセットは、製造ロボットを含む実製造シーンから作られ、ジェネレータは製造ロボットを含む製造シーンの2つのデータセット、すなわち、仮想データセット及び実データセット上で訓練されたCycleGAN実-仮想ネットワークである。本方法及び/又はプロセスはこの実施態様では2つのデータセット上で、CycleGANネットワークを訓練して、ドメイン適応ニューラルネットワークの訓練セットに近づけるために、テストシーンのテスト画像内のロボットの外観を適応させることを含み、この実施態様では、ドメイン適応学習方法による第3のドメインであり、仮想製造ロボットを含む仮想製造シーンから作られる。訓練の結果、実テスト画像を仮想化するように実テスト画像を修正することができる実-仮想ロボットジェネレータが得られる。

【 0 1 5 3 】

図22は、CycleGANジェネレータを用いて生成するステップS500の例を示す。この例では、試験データセットが各々1つ又は複数の製造ロボットを含む実製造シーンから作られる。図22は、テストデータセットの第1の現実シーン220及び第2の現実シーン222を示す。どちらのシーンも、多関節ロボットを含む製造シーンである。図22は、CycleGANジェネレータを適用することによって、第1のシーン220及び第2のシーン222からそれぞれ生成された第1の仮想シーン224及び第2の仮想シーン226をさらに示す。

【 0 1 5 4 】

さらに図4のフローチャートを参照すると、中間ドメインを決定するステップS50は、テストデータセットのシーンと生成された仮想シーンとを混合するステップS510を含むことができる。このような場合、混合は別のシーンをもたらす。

【 0 1 5 5 】

形式的には、

10

20

30

40

50

【数 4 2】

$$I$$

をテストデータセットのシーンとし、

【数 4 3】

$$I'$$

10

を生成された仮想シーンとし、

【数 4 4】

$$I''$$

を別のシーンとする。

【数 4 5】

$$I \text{ と } I' \text{ と}$$

20

を混合するステップ S 5 1 0 は、

【数 4 6】

$$I \text{ と } I' \text{ と}$$

を混合することができる任意の方法を指すことができ、その結果、混合

【数 4 7】

$$I''$$

30

が生じる。例では、テストデータセットのシーンと生成された仮想シーンとを混合するステップ S 5 1 0 が線形混合である。一実施形態では、線形混合が以下の式によって別のシーン

【数 4 8】

$$I''$$

40

を計算することを含む。

【数 4 9】

$$I'' = \alpha * I + (1 - \alpha) * I'.$$

【0 1 5 6】

図 2 3 は、線形混合の一例を示す図である。図 2 3 は、現実の製造ロボットを含む現実の製造シーン 2 3 0 と、現実のシーン 2 3 0 から生成された仮想シーン 2 3 2 とを示す。

50

図 2 3 はさらに、仮想製造シーン 2 3 4 である線形混合の結果を示す。

【 0 1 5 7 】

図 1 のフローチャートに戻って、中間ドメイン上で転送されたテストドメインのテストデータセットのシーンから空間的に再構成可能なオブジェクトを推論するステップ S 6 0 について説明する。

【 0 1 5 8 】

第 1 に、シーンから空間的に再構成可能なオブジェクトを推論するステップ S 6 0 は、1 つのシーンから空間的に再構成可能なオブジェクトを推論するステップと、シーンの入力データセットの各シーンについて、シーンの入力データセットの各シーンから空間的に再構成可能なオブジェクトを推論するステップとを含むことができる。特に、推論するステップ S 6 0 はドメイン適応型ニューラルネットワークのテスト段階を構成することができ、テストデータセットのすべてのシーンが入力された各シーン内の 1 つ又は複数の空間的に再構成可能なオブジェクトを推論するドメイン適応型ニューラルネットワークへの入力として（例えば、連続的に）提供される。追加的に又は代替的に、推論するステップ S 6 0 はドメイン適応学習方法及び / 又はドメイン適応推論方法及び / 又はプロセスの適用段階を構成することができ、入力データセットは、試験データセットに等しくないテストドメインのシーンのデータセットである。そのような場合、アプリケーション段階は、入力データセットを提供することと、入力データセットの各シーンをドメイン適応ニューラルネットワークに入力することと、各シーンについて推論（S 6 0）を実行することとを含むことができる。また、1 つのシーンにおける空間的に再構成可能なオブジェクトを推論するステップ S 6 0 は例えば、シーンにおける全ての空間的に再構成可能なオブジェクトを推論することによって、そのようなオブジェクトのうちの 1 つ以上を含むシーンにおける 1 つ以上の空間的に再構成可能なオブジェクトを（例えば、同時に又は連続して）推論するステップを含むことができることを理解されたい。

【 0 1 5 9 】

ここで、シーン内のより空間的に再構成可能なオブジェクトを推定するステップ S 6 0 の一部であってもよい、1 つのシーン内の 1 つの空間的に再構成可能なオブジェクトを推定するステップ S 6 0 について説明する。ドメイン適応ニューラルネットワークを適用することによってシーン内の空間的に再構成可能なオブジェクトを推論するステップ S 6 0 は、シーンをドメイン適応ニューラルネットワークに入力することにある。次に、ドメイン適応型ニューラルネットワークは、空間的に再構成可能なオブジェクトの位置に関するデータを出力する。例えば、ニューラルネットワークはシーンの（例えば、デジタル）画像表現を出力することができ、それぞれの境界ボックスは、推論された空間的に再構成可能なオブジェクトを包含する。推論はまた、そのような境界ボックスの計算、及び / 又は境界ボックスのラベル付け（例えば、シーンの空間的に再構成可能なオブジェクトを包含するものとして）、及び / 又は推論された空間的に再構成可能なオブジェクトのラベル付け（空間的に再構成可能なオブジェクトとして）を含んでもよい。ニューラルネットワークの出力は、データベースに記憶され、及び / 又はコンピュータのスクリーン上に表示されてもよい。

【 0 1 6 0 】

図 2 4 ~ 2 9 は、実製造シーンにおける製造ロボットの推論を示す。図 2 4 ~ 図 2 9 は、それぞれ、ドメイン適応ニューラルネットワークへの入力として供給されるテストドメインの対応する実シーンと、対応するシーンに含まれるすべての製造ロボットの周りで計算されアノテーション付けされた境界ボックスとを示す。図 2 6 ~ 図 3 0 に示すように、現実の場面において、少なくとも部分的に遮蔽されたいくつかのロボット及び / 又はロボットを推測することができる。

【 0 1 6 1 】

空間的に再構成可能なオブジェクトの推論は、実際には認識（又は分類）のタスクと、テストデータセットの実シーンにおける空間的に再構成可能なオブジェクトの位置特定のタスクとを共同で実行するドメイン適応型ニューラルネットワークの少なくとも一部を形

10

20

30

40

50

成する検出器によって実行されてもよい。位置特定のタスクは、検出器に入力されたシーンの空間的に再構成可能なオブジェクトを各々が包含する境界ボックスを計算することを含む。前述のように、境界ボックスは軸が画像の側に平行である矩形ボックスであり、4つの座標によって特徴付けられる。例では、検出器が検出器に入力されたシーンの各空間的に再構成可能なオブジェクトについて、適切な比率及びスケールでオブジェクトを中心とする境界ボックスを返す。分類のタスクは計算された各境界ボックスに対応するラベル（又はアノテーション）をラベル付けし、ラベルに信頼スコアを関連付けることを含む。信頼性スコアは、境界ボックスが入力として提供されるシーンの空間的に再構成可能なオブジェクトを真に包含するという検出器の信頼性を反映する。信頼度スコアは、0と1との間の実数であってもよい。このような場合、信頼スコアが1に近ければ近いほど、検出器は、対応する境界ボックスに関連付けられたラベルが空間的に再構成可能なオブジェクトに真にアノテーションを付けることについての信頼度が高くなる。

10

【0162】

ドメイン適応ニューラルネットワークが現実のシーンにおけるオブジェクトの推論のために構成される例では例えば、テストドメインが現実のシーンから作られ、中間ドメインが仮想シーンから作られる場合、ドメイン適応ニューラルネットワークはエクストラクタをさらに備えてもよい。エクストラクタは現実のシーンの画像表現を出力し、それらをドメイン適応ニューラルネットワークの他の部分、例えば、検出器又は検出器の一部に供給し、例えば、空間的に再構成可能なオブジェクトの検出を監視するように構成されたドメイン適応ニューラルネットワークの一部を形成する。例では、この手段がエクストラクタが中間ドメインのシーンを処理して、例えば、実データスタイルのためのロバストな畳み込みフィルタを学習することによって、シーンの画像表現を出力することができることである。エクストラクタのニューロンの層は、典型的には出力されるべき現実の又はフォトリアリスティックな画像の表現を符号化する。これは、エクストラクタが中間ドメインに属するシーンの画像表現の出力を監視し、一方、ドメイン適応ニューラルネットワークの少なくとも1つの他の部分（例えば、前述の検出器）が第3のドメインの各入力シーンにおける空間的に再構成可能なオブジェクトの検出（例えば、局所化及び分類）を監視することを意味する。前記検出は、生徒エクストラクタによって出力された前記画像表現に基づくことができる。これにより、実データを効率的に処理して出力することができる。

20

【0163】

ドメイン適応ニューラルネットワークへの入力として提供される各シーンはテストドメインに属するが、決定された中間ドメイン上で転送されるシーンである。言い換えれば、シーンを入力としてドメイン適応ニューラルネットワークに提供することは、中間ドメイン上のシーンを転送することと、転送されたシーンを入力としてドメイン適応ニューラルネットワークに供給することとを含む。シーンの転送は、1つのドメインから別のドメインにシーンを転送することができる任意の方法によって実行することができる。例では、シーンを転送することはシーンを別のシーンに変換することと、任意選択で、シーンを変換されたシーンと混合することとを含むことができる。そのような場合、変換及び任意選択で混合は、図4のフローチャートを参照して前述した例のように実行することができる。特に、仲介ドメインの場面は、実際には仲介ドメイン上で転送されたテストデータセットの場面である可能性がある。

30

40

【0164】

次に、プロセスの実施について説明する。

【0165】

この実装は製造データ不足問題に対処するために、仮想世界における学習フレームワークを提案する。この実装は特に、学習のための仮想シミュレーションのデータセットと、画像処理方法、すなわち深部ニューラルネットワーク（DNN）の効率とを利用して、ドメイン適応ニューラルネットワークを学習する。

【0166】

仮想シミュレーションは、訓練のための仮想データを取得する1つの形成を構成する。

50

後者は、機械学習タスクのために必要とされる訓練データを得ることの困難さに対処するための有望な技術として現れてきた。仮想データは実データの特性を模倣するようにアルゴリズム的に生成されるが、自動的で、エラーがなく、費用のかからないラベリングを有する。

【0167】

異なる製造装置の作業メカニズムをより良く探索するために、この実装により仮想シミュレーションを提案した。形式的には、異なる製造ツールがそれらが機能するときに提示することができる異なる状態を介して、学習プロセスにおいて考慮される。

【0168】

深層ニューラルネットワーク：ニューラルネットワークにおいて学習するための強力な技術のセットであり[19]、これは、コンピュータが観察データから学習することを可能にする生物学的にインスパイアされたプログラミングパラダイムである。画像認識において、DNNの成功は他の画像分類法(SVM、Boosting、Random Forestなど)で使われる手で設計された低レベル特徴(Zernikeモーメント、HOG、Bag of Words、SIFTなど)とは対照的に、豊富な中レベルメディア表現を学習する能力に起因する。より具体的には、DNNが生データに基づくエンドツーエンド学習に焦点を当てている。言い換えれば、それらは、生の特徴で始まりラベルで終わるエンドツーエンド最適化を達成することによって、可能な最大限まで特徴エンジニアリングから離れる。

【0169】

ドメイン適応：機械学習及び伝達学習に関連するフィールドである。このシナリオはソースデータ分布から、異なる(しかし関連する)ターゲットデータ分布に対して良好にパフォーマンスを示すモデルを学習することを目的とする場合に生じる。ドメイン適応の目的は、知識を1つのドメインから別のドメインに移転又は適応させるための効果的なメカニズムを見つけることである。これは、列車と試験データとの間のクロスドメインギャップを改善するためにクロスドメインブリッジングコンポーネントを導入することによって達成される。

【0170】

提案された実施は特に、プロセスのオフライン及びオンライン段階を示す図31によって示される。

【0171】

1. オフライン段階：この段階は、現実世界で推論する際に有能であると考えられる仮想データを使用してモデルを訓練することを目的とする。それは、2つの主要なステップを含む。この段階はユーザからは透過的であることに注意する。

【0172】

(1) 仮想製造環境における仮想シミュレーションにより、自動的にアノテーション付けされる代表的な仮想訓練データを生成する。仮想製造環境は幾何学的に現実的であるが、フォトリアリスティックではない。アノテーションの種類は、ターゲットの推論タスクによって異なる。

【0173】

(2) ニューラルネットワークモデルは、仮想データに基づいて学習される。これは、ドメイン適応DNNベースのモデルで構成される。

【0174】

2. オンライン段階：現実世界のメディアが与えられると、学習されたドメイン適応モデルが、製造装置を推論するために適用される。

【0175】

このプロセスの実施は特に、完全に監視された物体検出の分野に関する。具体的には、この具体的な実装において、著者らは製造コンテキストに関心を持ち、目標は工場画像における産業用ロボットアームを検出することである。前記完全に監督された技術分野について、訓練データセットは、訓練のための仮想データの使用を正当化する、数十万までの

10

20

30

40

50

アノテーション付きデータを含むべきである。

【0176】

最先端の物体検出器は深層学習モデルに基づいている。手動で値を設定することができない何百万ものパラメータが、これらのモデルを特徴付ける。したがって、これらのパラメータは、学習アルゴリズムのおかげで設定されなければならない。学習アルゴリズムがモデルパラメータを更新しているとき、モデルは「訓練モード」にあると言われる。各入力に関連付けられたアノテーションのおかげで、入力ごとにモデルを連続的に「修正」することからなる。アノテーションはモデルの出力が真か偽かを評価することを可能にする特定の入力に関連付けられたデータのセットである。例えば、ネコとイヌの画像を区別するために訓練されたオブジェクト分類器はネコとイヌのアノテーションされた画像のデータセットを必要とし、各アノテーションは「ネコ」又は「イヌ」である。したがって、その訓練モードにおいて、オブジェクト分類器が入力中のネコ画像について「犬」を出力する場合、学習アルゴリズムは、そのパラメータを更新することによってモデルを修正する。アノテーションされたデータセットのおかげでモデルの訓練を監督する方法は「監督学習」と呼ばれる。モデルが訓練されると、そのパラメータの更新を停止する。その後、モデルは新しい入力（すなわち、訓練モード中の未知の入力）を処理し、検出結果を戻すためにのみ使用される。これは「テストモード」と呼ばれる。

10

【0177】

本実施形態のコンテキストでは、物体検出器が認識（又は分類）のタスクと位置特定のタスクとを共同で実行する手段に、「検出」のタスクとして2つの異なった出力を返す。

20

【0178】

1. 局在化出力：オブジェクト局在化は、境界ボックスのおかげで行われる。境界ボックスは、軸が画像の側に平行である矩形ボックスである。それは、4つの座標によって特徴付けられる。理想的には、物体検出器が各物体に対して、適切な比率及びスケールで、物体の中心に境界ボックスを戻す。

【0179】

2. 分類出力：オブジェクト分類は、各境界ボックスの信頼スコアに関連付けられたカテゴリラベルのおかげで行われる。信頼スコアは、0と1との間の現実である。それが1に近ければ近いほど、オブジェクト検出器は、対応する境界ボックスに関連するカテゴリラベルについてより確信がある。

30

【0180】

この実装は特に、オフライン及びオンラインの両方の段階で使用されるドメイン適応成分を示す図32によって示される。

【0181】

本発明は、生成された仮想データを用いて物体検出モデルを学習することに関する。オフラインとオンラインの両方の段階内の仮想 - 実ドメインギャップに対処するために、いくつかのドメイン適応アプローチを適用した。

【0182】

1. オフライン段階：この段階は、現実世界で推論する際に有能であると考えられる仮想データを使用してモデルを訓練することを目的とする。この段階は、2つの主要なステップを含む。

40

【0183】

(1) 仮想製造環境における仮想シミュレーションは、境界ボックス及び「ロボット」ラベルで自動的にアノテーション付けされる代表的な仮想訓練データを生成する。

【0184】

(2) ドメイン適応変換

(i) 訓練セットレベルでは、後者がデータ分布に関して現実世界により近い新しい訓練ドメインに変換される。

(ii) モデルアーキテクチャレベルでは、ドメイン適応モジュールが問題のある仮想 - 実ドメインシフトに対処するためにモデルに挿入される。

50

【 0 1 8 5 】

２．オンライン段階：実世界のメディアは、学習された検出モデルによってより良好に処理される新しいテストドメインに変換された後、後者によって推論される。言い換えれば、新しいテストドメインは、元の実ドメインよりも訓練ドメインに近い。

【 0 1 8 6 】

この実施態様の１つの特殊性は、仮想 - 実画像変換精度の課題に遭遇しないことであることに留意することが重要である。現在の研究の背後にある直感は仮想データを現実世界に移行しようとする代わりに、仮想世界と現実世界との間の中間ドメインを考慮することである。形式的には、仮想訓練データ及び実テストデータが、ドメインシフトが最小である訓練及びテスト中間ドメインにそれぞれドラッグされる。

10

【 0 1 8 7 】

上述のフレームワークに照らして、現在論じられている実施は２つの主要なステップに基づくことを理解することができる。

- ・学習のための仮想データ生成。
- ・ドメイン適応。

【 0 1 8 8 】

工場イメージの不足に対処するために、この実装は、シミュレーションに基づく仮想データ生成アプローチに基づく。訓練セットは、３Ｄレンダリングされた工場シーンからなる。これらの工場シーンは、画素レベルで関心対象（ロボット）をセグメント化するマスクを伴う会社の内部設計アプリケーションによって生成される。各シーンは、ロボットが有するのであろう複数の機能状態を記述する複数のフレームにシミュレートされる。

20

【 0 1 8 9 】

このステップは製造状況においてロボットを検出するタスクのために、特徴的で複雑な特徴を有する仮想画像セットを生成する。それは「単純である」単一物体検出タスクのように見えるが、ベースライン検出器及び最新技術のドメイン適応ネットワークはこの問題において能力がない。形式的には、生成されたセットがフォトリアリスティックではなく、むしろ「漫画的」である。生成された仮想シーンの単純さは我々の訓練データを生成する容易さを反映しており、これは本実施形態の主要な利点である。同時に、関心対象の複雑さは、この種の仮想データに取り組む最新のアプローチ、すなわちドメインランダム化DR技術を使用することを想定することを困難にする（[7 , 1 5]を参照）。

30

【 0 1 9 0 】

実際、ドメインランダム化は、非フォトリアリスティックな方法で環境をランダムに摂動させることによってフォトリアリズムを放棄する。このDRの背後にある直感はネットワークに、画像の本質的な特徴に焦点を当て、関心のないシーン内のオブジェクトを無視するように学習させるアルゴリズムを教示することである。より形式的には、仮想データに欠ける主要な現実的な特徴が画像の低レベルの形態、例えば、テクスチャ、照明、及びシェーディングである。DRはアルゴリズムがオブジェクトを認識するためにそれらに依存しないように、訓練データセットにわたるこれらの特徴の摂動を保証する。同じロジックに従うと、DRは実データと仮想データ間の共通機能に影響しない。使用される仮想データ生成の範囲では、学習モデルが学習に頼ることが期待される正しい特徴が形状及び輪郭であることを理解することができる。これは、DR技術において重要である。

40

【 0 1 9 1 】

さらに、ロボットである本実施の対象は、連続した値の範囲によって較正される多くのジョイントの構成（一般に６自由度）を有する。これにより、各ロボットタイプに可能な状態の無限範囲が与えられる。したがって、学習のための形状及び輪郭を除くすべての特徴をランダム化することは、この種の複雑な産業用物体には最適ではない。実際には、物体の形状及び輪郭を表すために必要とされる変動性が、物体が複雑である場合には満たすことが困難である。特に、ロボットの例では、画像中の前記物体を認識するために、無限の関節の位置を持つ様々なロボットタイプの特徴的な形状や輪郭を抽出し、焦点を合わせることができるモデルを学習することは実際には考えられない。

50

【0192】

提案した実装は、この場合のようにデータが複雑な場合に有能な学習問題を構築する。

【0193】

提案したドメイン適応アプローチは、部分ドメインランダム化技術と考えることができる。形式的には、モデルが学習している正しい特徴を、データ特徴をランダム化するか、又は仮想のものよりも現実世界をより良く表す新しいドメインに変換することによって、増強することを提案する。例えば、テキストチャ特性の特異性をよりよく探索することができる。そうするために、3つの新規な技術が、現在議論されている実施において実施される：

【0194】

1. 仮想から実への変換

クロスドメインシフトを改善するために、最新技術の論文における一般的な方法は仮想データがあたかも現実のドメインから引き出されたかのように見えるように仮想データの外観を変更することであり、すなわち、仮想データは、現実の世界をより良く表す分布を獲得する。そのために、仮想画像は類似オブジェクト間の対応関係なしに、現実シーンを使用して全体として修正される。しかしながら、画像から画像への変換の最適な性能を得るためには、実ロボットテキストチャを仮想シーン内のロボットに転送させ、背景については同じものを転送させることが有益であると推測することができる。

現在の実施は、訓練セットからの各媒体内のロボット及び工場の背景に別々に異なる画像対画像変換ネットワークを適用することによって、この目標とされた外観転送に基づいて構築される。

【0195】

この目的のために、CycleGANモデル[20]を使用して、画像を1つのドメインから別のドメインに変換する。本発明は1つの画像セットの特別な特徴をキャプチャし、これらの特徴を第2の画像コレクションに変換する方法を、すべて対になった訓練例がない場合に、理解することができる方法を提案する。1対1マッピングを有するこの緩和は、実際にはこの定式化を非常に強力なものにする。これは生成モデル、具体的にはCycleGANと呼ばれる生成アドバーサルネットワーク(GAN)によって達成される。

【0196】

上述のように、2つの異なるモデルは、別々に訓練される。

・CycleGAN現実-実ネットワーク1：2つの工場背景セット、すなわち現実セット及び実セットで訓練される。その結果、仮想画像内の背景を修正するために使用される仮想-実工場ジェネレータが得られる。

CycleGAN仮想-実 2：2つのロボットセット、すなわち、仮想セット及び実セットで訓練される。その結果、仮想画像内のロボットを修正するために使用される仮想-実ロボットジェネレータが得られる。

【0197】

2. グラムベースの蒸留

ドメインシフトに関してモデルのロバスト性をより良く強調するために、検出器レベルで実装した新しい知識蒸留アプローチを提案した。最新技術の検出器の大部分([18, 9]参照)は、畳み込み層の深層ネットワークである特徴エクストラクタを含むことに留意されたい。特徴エクストラクタは、検出器の残りの部分に供給されるべき画像表現を出力する。著者らの蒸留の背後にある直感は実データスタイルのためのロバストな畳み込みフィルタを学習するために検出モデルをガイドすることであり、これはモデルをデータ外観に対して不変にする主目的に整合させている。

【0198】

形式的には、この実施が検出器を仮想媒体上で訓練し、同時に、検出器の特徴エクストラクタの重みを更新して、我々の検出器とオープンソースからの実データ上で予め訓練された第2の検出器の両方に供給される実画像の2つのレイヤ表現によって符号化されたスタイルに近づけることからなる。

10

20

30

40

50

【 0 1 9 9 】

ある層

【数 5 0 】

 ℓ

における画像のスタイル表現は、この層における特徴分布によって符号化されることに留意されたい。この実装のために、特徴分布埋め込みの一形態である特徴相関を符号化するグラム行列を使用する。形式的には長さ

【数 5 1 】

10

 N^ℓ

のフィルタを有する応答

【数 5 2 】

 F^ℓ

の層

【数 5 3 】

20

 ℓ

について、特徴相関はグラム行列

【数 5 4 】

 $G^\ell \in R^{N^\ell \times N^\ell}$

によって与えられ、ここで、

30

【数 5 5 】

 $G_{i,j}^\ell$

は層

【数 5 6 】

 ℓ

40

内のベクトル化された特徴マップ i と j との間の内積である。

【数 5 7 】

 $G_{i,j}^\ell = \sum_k F_{i,k}^\ell F_{j,k}^\ell$

【 0 2 0 0 】

提案された蒸留の背後にある直感は学習中に、検出器の特徴エクストラクタの重みが更新され、その結果、検出器が、実データ上で互いに事前訓練された特徴エクストラクタが生成するのであろうものに近いデータスタイル表現を出力することを保証することである

50

。そうするために、現在論じられている実施は、特徴エクストラクタレベルでグラムベースの蒸留損失

【数 5 8】

$$L_{distillation}$$

を実施する。

【 0 2 0 1】

実画像（これを

【数 5 9】

$$F^\ell \text{ 及び } E^\ell$$

10

という）が与えられると、対応する特徴マップは、検出器の特徴エクストラクタ及び事前訓練された特徴エクストラクタをある層

【数 6 0】

$$\ell$$

20

において出力する。したがって、層

【数 6 1】

$$\ell$$

における提案された蒸留損失は、対応するグラム行列

【数 6 2】

$$G_F^\ell \text{ と } G_E^\ell \text{ と}$$

30

の間のユークリッド距離

【数 6 3】

$$L_{distillation}^\ell = \|G_F^\ell - G_E^\ell\|$$

である。

層

【数 6 4】

$$\{\ell_1, \dots, \ell_m\}$$

40

から知識を蒸留することを考慮すると、蒸留損失は次のように定義される。

【数 6 5】

$$L_{distillation} = \sum_{1 \leq i \leq m} L_{distillation}^{\ell_i}$$

50

訓練損失の合計は

【数 6 6】

$$L_{training} = L_{detector} + L_{distillation}$$

である。

【 0 2 0 2】

3．実から仮想への変換

仮想画像を現実のドメインに変換するタスクが完全には達成され得ないことが予想される。この手段は、修正されたデータが現実の分布に完全にマッチすることなく、仮想データよりも現実の世界に近いということである。そのため、修正された仮想データは、合成ドメインと実ドメインとの間に位置する新しいドメイン

【数 6 7】

$$D_1$$

に転送されると仮定する。この研究の 1 つの貢献はこの後者を別の新しいドメイン

【数 6 8】

$$D_2$$

にドラッグすることにより、ドメイン適応を実テストデータに拡張することである。このアプローチを使用するためには分布に関して

【数 6 9】

$$D_2$$

は実ドメインよりも

【数 7 0】

$$D_1$$

に近くなければならず、これはデータ表現を比較することによって本出願の場合に確認される仮定である。この形態の適応は、テストデータに実データから仮想データへの変換、すなわち訓練フェーズで仮想データに以前に適用された逆演算を適用することによって達成される。

【 0 2 0 3】

本実施形態のコンテキストでは、変換が 2 つのロボットセット、すなわち、仮想セット及び実セット上で訓練された CycleGAN 実 - 仮想ネットワークを訓練することによって保証される。この結果、実テスト画像を修正するために使用される実 - 仮想ロボットジェネレータが得られる。このレベルでは、著者らが検出器によって認識されるために、テスト画像におけるロボットの外観の適応に主に興味があるので、ロボットセット上で CycleGAN ネットワークを訓練することを選択した。また、訓練セットに対して行われたように、ドメインを別々にターゲット化することによって、テストデータに対してより高度な画像変換を実行することも可能であることに留意されたい。しかし、改善は限られており、テスト画像全体に 1 つの CycleGAN ネットワークを適用することに満足している。

10

20

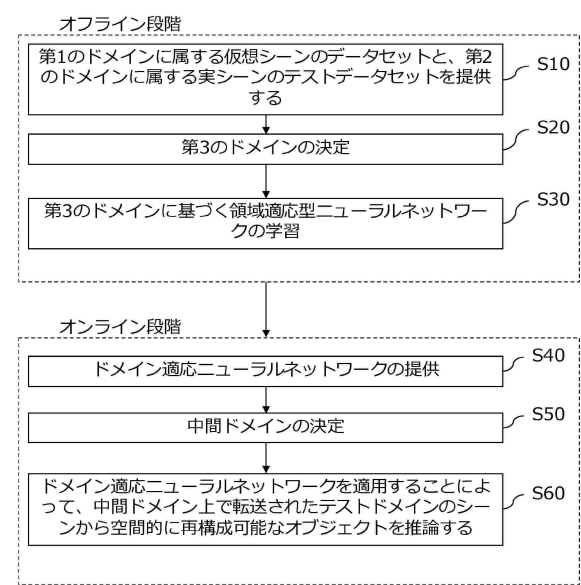
30

40

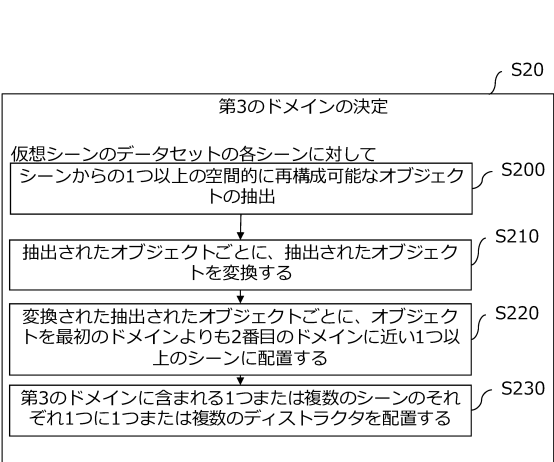
50

【 図 面 】

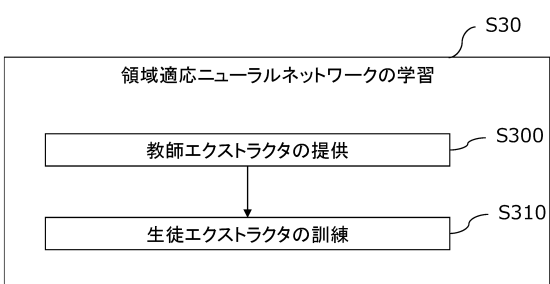
【 図 1 】



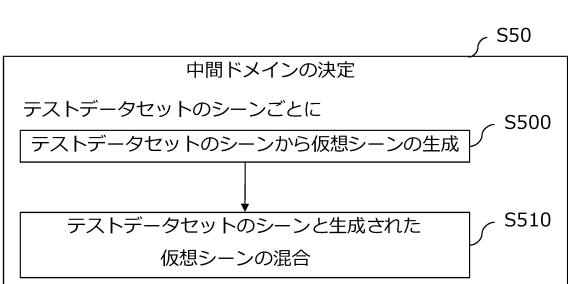
【 図 2 】



【 図 3 】



【 図 4 】



10

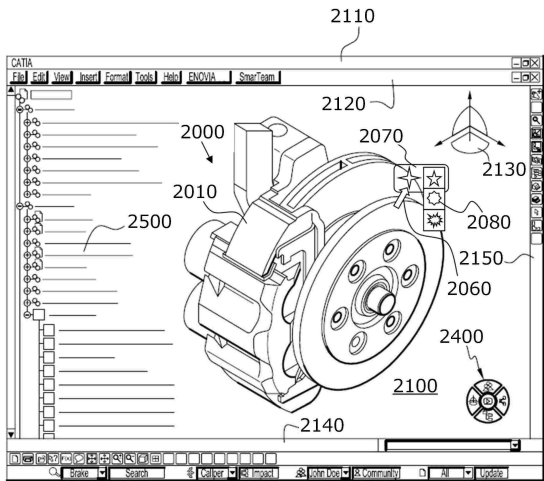
20

30

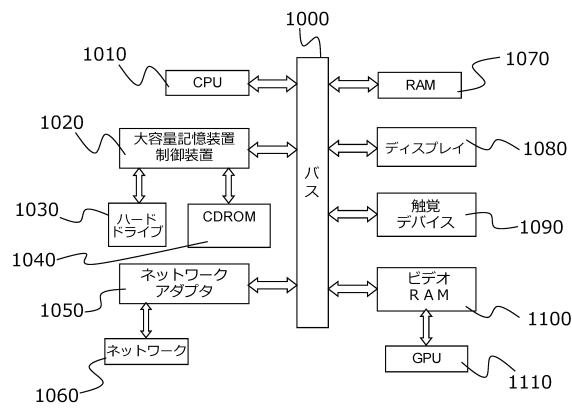
40

50

【図 5】



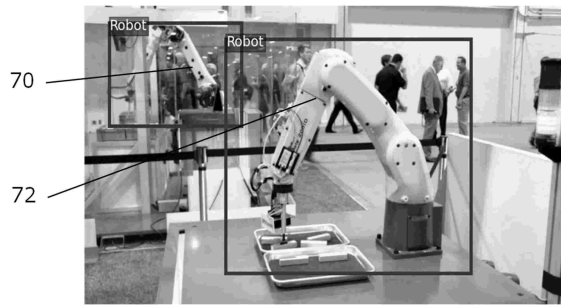
【図 6】



10

20

【図 7】



【図 8】



30

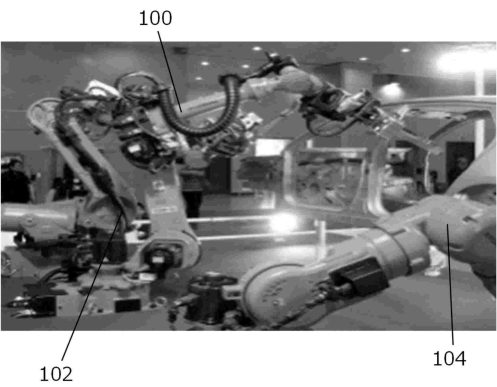
40

50

【 図 9 】



【 図 1 0 】

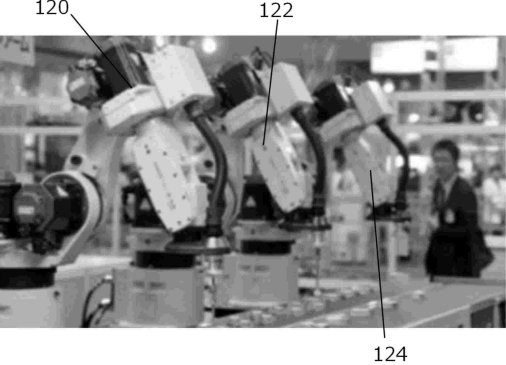


10

【 図 1 1 】



【 図 1 2 】



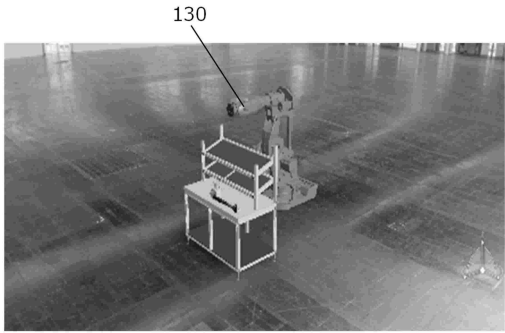
20

30

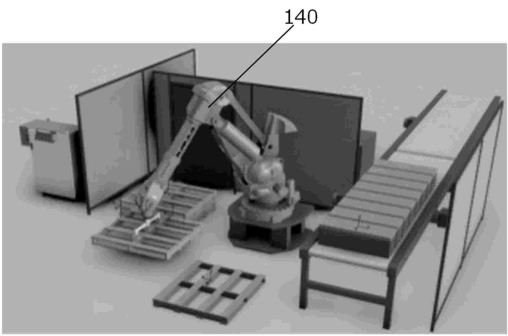
40

50

【 図 1 3 】

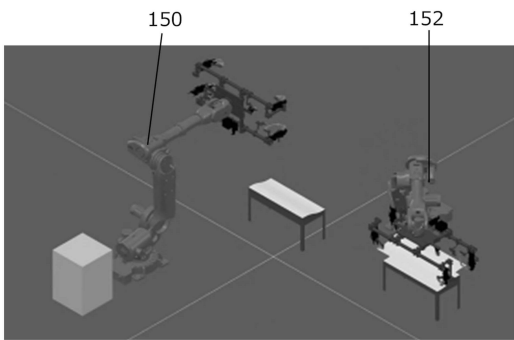


【 図 1 4 】

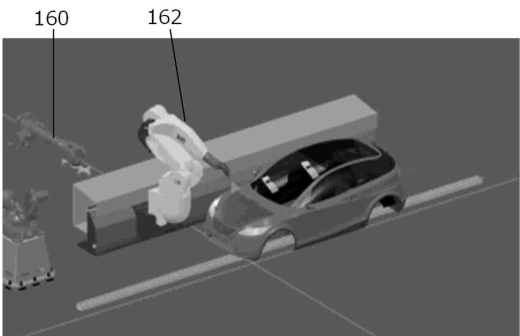


10

【 図 1 5 】



【 図 1 6 】



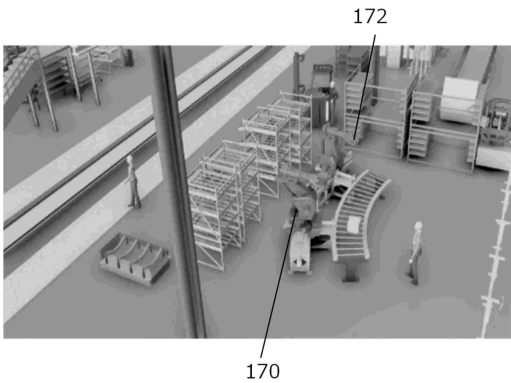
20

30

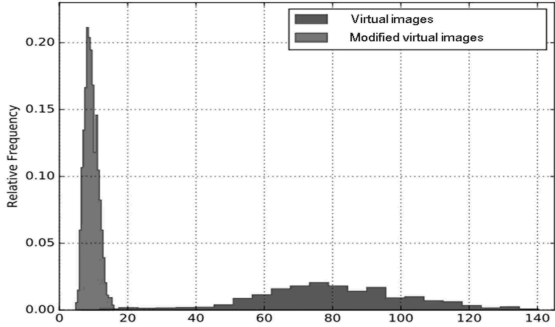
40

50

【図 17】

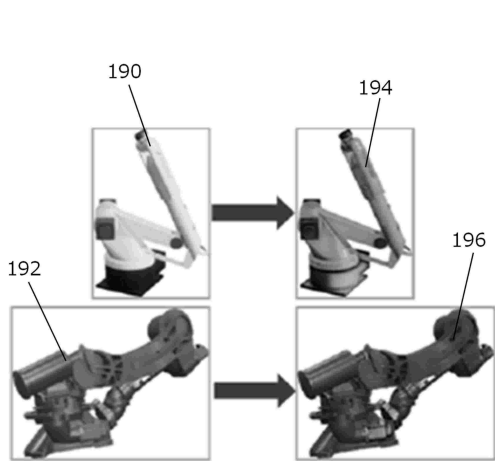


【図 18】



10

【図 19】



【図 20】



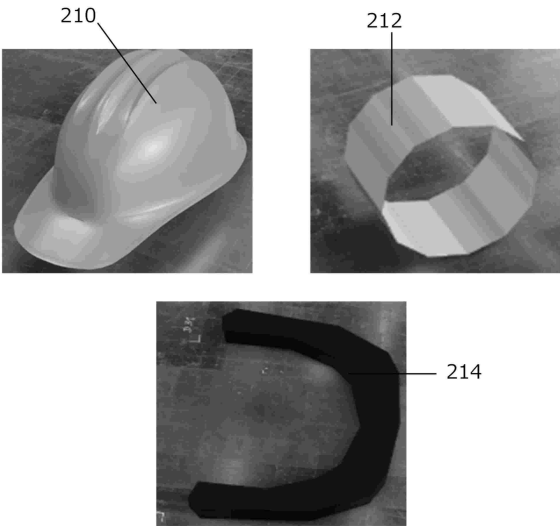
20

30

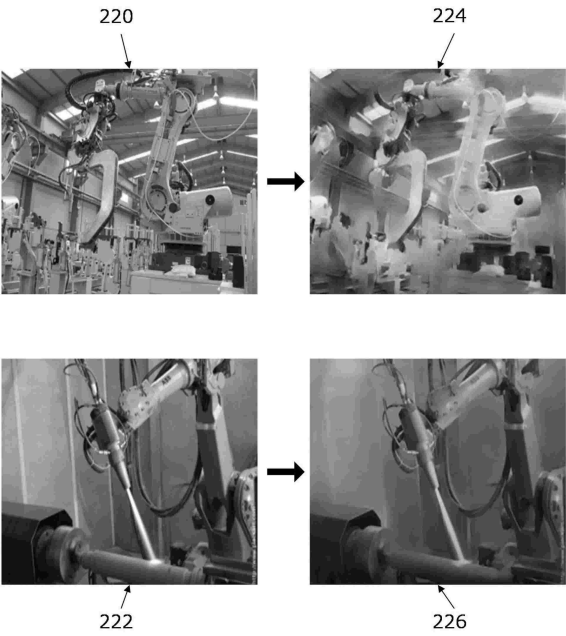
40

50

【 図 2 1 】



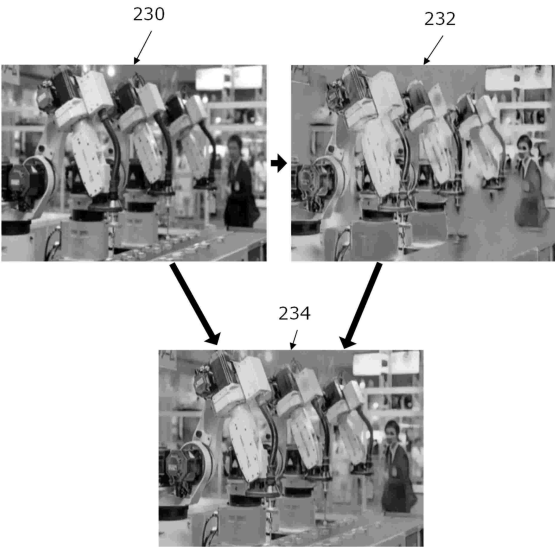
【 図 2 2 】



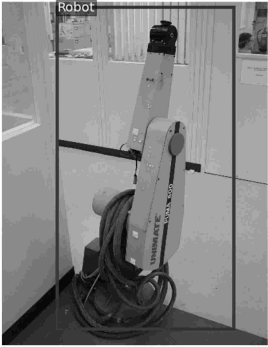
10

20

【 図 2 3 】



【 図 2 4 】



30

40

50

【 図 2 5 】

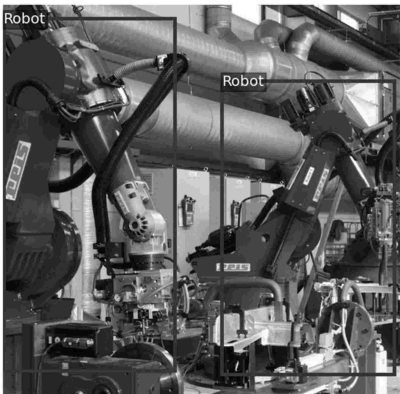


【 図 2 6 】



10

【 図 2 7 】



【 図 2 8 】



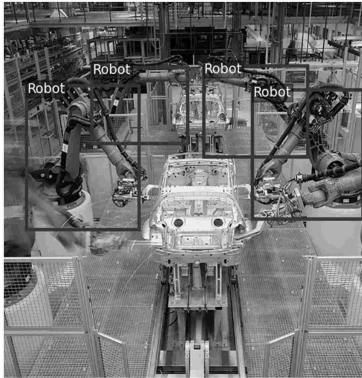
20

30

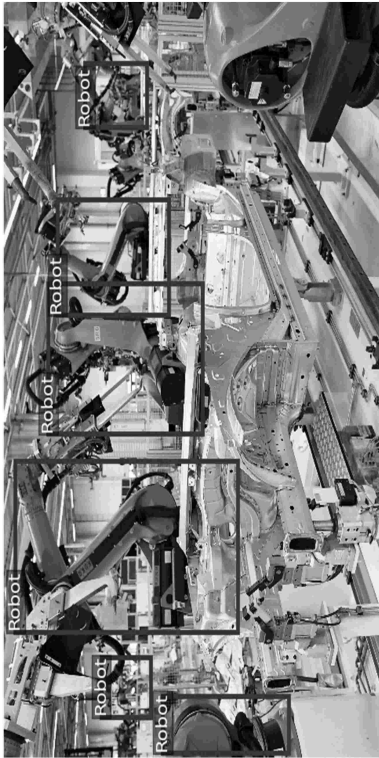
40

50

【図 29】



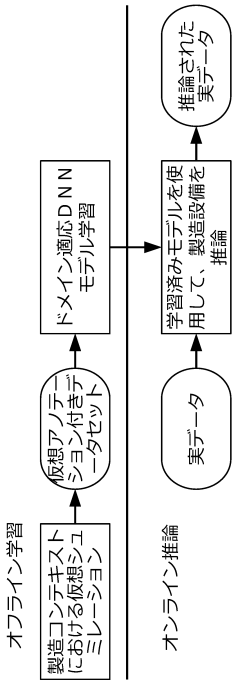
【図 30】



10

20

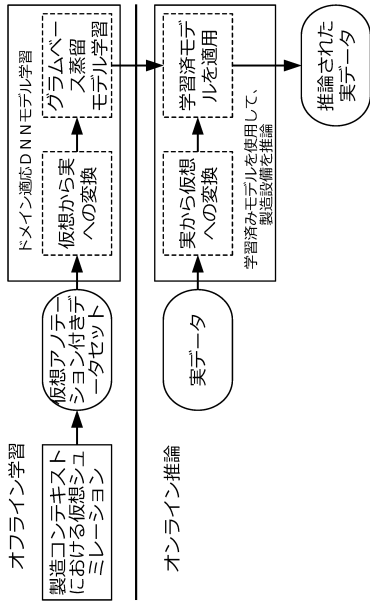
【図 31】



30

40

【図 32】



50

フロントページの続き

- (72)発明者

メズガニ マリエム
フランス国 7 8 1 4 0 ヴェリジーヴィラクブレー マルセルダッソー通り 1 0 ダッソーシステムズ社内
- 審査官

広 島 明芳
- (56)参考文献

米国特許出願公開第 2 0 1 8 / 0 3 4 5 4 9 6 (U S , A 1)
特開 2 0 2 0 - 1 1 9 5 5 3 (J P , A)
Jingwei Zhang, et al. , VR-Goggles for Robots: Real-to-Sim Domain Adaptation for Visual Control , IEEE Robotics and Automation Letters , 米国 , 2019年01月20日 , <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8620258>
- (58)調査した分野

(Int.Cl. , D B 名)
G 0 6 T 7 / 0 0 - 7 / 9 0
G 0 6 N 3 / 0 2
G 0 6 V 1 0 / 0 0 - 2 0 / 9 0