When Deep Learning Meets Data Alignment: A Review on Deep Registration Networks (DRNs)

Victor Villena-Martinez [1,*], Sergiu Oprea [1], Marcelo Saval-Calvo [1,*], Jorge Azorin-Lopez [1], Andres Fuster-Guillo [1] and Robert B. Fisher [2]

[1] Department of Computer Technology, University of Alicante, 03690 Alicante, Spain; soprea@dtic.ua.es (S.O.); jazorin@dtic.ua.es (J.A.L.); fuster@dtic.ua.es (A.F.G.)
[2] School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK; rbf@inf.ed.ac.uk
* Correspondence: vvillena@dtic.ua.es (V.V.M.); msaval@dtic.ua.es (M.S.C.)

Abstract: This paper reviews recent deep learning-based registration methods. Registration is the process that computes the transformation that aligns datasets, and the accuracy of the result depends on multiple factors. The most significant factors are the size of input data; the presence of noise, outliers and occlusions; the quality of the extracted features; real-time requirements; and the type of transformation, especially those defined by multiple parameters, such as non-rigid deformations. Deep Registration Networks (DRNs) are those architectures trying to solve the alignment task using a learning algorithm. In this review, we classify these methods according to a proposed framework based on the traditional registration pipeline. This pipeline consists of four steps: target selection, feature extraction, feature matching, and transform computation for the alignment. This new paradigm introduces a higher-level understanding of registration, which makes explicit the challenging problems of traditional approaches. The main contribution of this work is to provide a comprehensive starting point to address registration problems from a learning-based perspective and to understand the new range of possibilities.

Keywords: registration; 3D alignment; neural networks; Deep Registration Networks

1. Introduction

In the context of computer vision, registration is the process of aligning data into a common frame of reference. In other words, it aligns datasets—captured from different sources, viewpoints, and/or at a different time step—by means of geometric transformations. Here, we consider both 2D and 3D data, and data in point sets, grids, and meshes. Rigid and non-rigid registration has already been widely addressed in the computer vision literature through potential applications mostly for data analysis such as body modeling [1] for pose analysis; computed tomography registration [2,3] for medical diagnosis; multi-camera registration for robot guidance [4]; and applications in object classification on assembly lines [5], among others. In the aforementioned applications, registration represents a crucial component. It fuses a vast amount of raw data captured under different scenarios, which greatly facilitates the analysis process.

The growing number of available consumer-grade devices, such as RGB-D cameras and LiDAR sensors, provides quasi-unlimited and cheap data of different modalities. However, the raw data must be previously structured, either hierarchically or semantically, to extract high-level information. The vast amount of data has surpassed the potential of the traditional registration paradigm and led researchers to consider learning-based approaches. Dealing with a huge amount of raw and unstructured multidimensional data is not straightforward, yet they satisfy Deep Learning (DL) methods that are known to be data-hungry. Learning-based approaches have proliferated in recent years, making great strides in different fields [6,7,101]. Considering this success, deep learning-based registration approaches are poised to leap over the previous state of the art in the registration paradigm. However, existing DL-based techniques for rigid and non-rigid registration, mostly in an n-dimensional

space, are far from accurate and reliable. Furthermore, the direct application of DL techniques to the problem of registration is not straightforward; its lack of maturity and the rapid state of this field make it difficult to keep up with the latest trends and track them properly.

1.1. Review Scope

This paper reviews state-of-the-art learning-based approaches to registration. The ability of deep neural networks to generalize from training data and manage geometric properties has created a new subfield at the intersection between learning and registration algorithms. Although some reviews of registration have been performed [8–10], no reviews address learning-based approaches for registration without focusing on a specific scope such as medical image registration or image localization.

The contributions of this paper are: (1) we provide a global overview of learning-based registration methods by proposing a well-defined framework that encompasses both the traditional and learning-based approaches; and (2) we review the recent learning-based registration approaches, which have been classified according to a proposed taxonomy to foster discussion.

Figure 1 graphically summarizes the scope of this paper. The figure shows at the top the four main stages of the traditional pipeline for registering two given inputs (P and Q). These stages are: Target Selection (yellow, oblique lines), which defines the fixed input that the other input is going to be aligned to; Feature Extraction (red, dotted pattern), which computes the set of features ω and φ for each input; Feature Matching (green, squared lines) to find correspondences between the previously extracted features; and Pose Optimization (blue, vertical lines), which is the process to minimize the distance error between both inputs. The right end is the final result transformation ($[R, t]$), which is the rotation R and translation t parameters that indicate how data should be transformed to be aligned. These stages and terminology are further explained in Section 2.

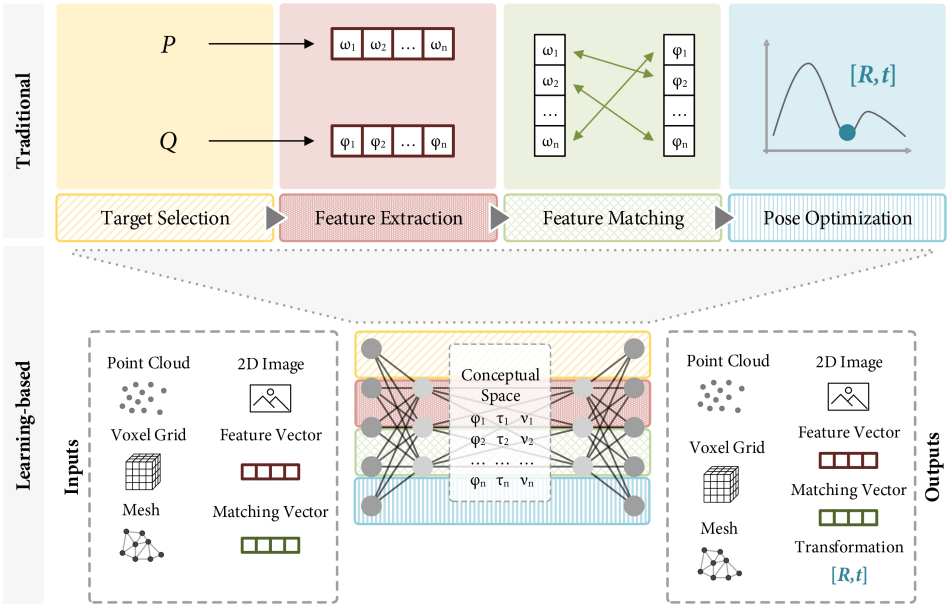

Figure 1. Registration framework. Schematic of the registration process in the traditional pipeline on the top part, and the learning-based approaches reviewed in this paper at the bottom. Learning-based approaches encapsulate the traditional pipeline in a conceptual space, allowing different types of inputs depending on the data and the number of encapsulated steps. The conceptual space is the learned parameters during the training process, and theoretically could also be considered as an input for the registration process.

Vertically, the schematic is divided at the top with the traditional stages clearly defined and at the bottom the learning-based approaches that are reviewed here, represented by a neural network, and the possible data for inputs and outputs. Having in mind this information, the inputs for the

learning-based proposal could be in different formats, such as point clouds, voxel grids, meshes, etc. In addition to full end-to-end approaches, some methods accept as inputs the result of the Feature Extraction or Matching stages in the traditional pipeline. The output could be some data in a specific format as well as the result of the stages from the traditional pipeline (feature vector, matching vector, and transformation). Nevertheless, with the new learning-based approaches, a new conceptual kind of space appears, containing learned properties about the object, materials, and their behavior that can be registered with the input data (e.g., aligning input data of an inflated ball with its deflated state restricted by the physical behavior of the object, rather than aligning with a final deflated target). The conceptual space is modeled by a neural network and its training process, and, theoretically, it could be considered as an input to the registration process, but it is not an input that one might use in every registration instance since it is an internal representation. This allows the network to encode conceptual models such as physical phenomena (e.g., force vector and symbolic/conceptual information such as "sporty, comfy") or mathematical rules. Besides, the neural network could perform one or more phases from the traditional pipeline (represented by the colored rectangles (see Figure 1) shown in the network). Considering all of this, we can see that there are multiple possibilities, combining inputs from different stages and outputs from the learning-based approaches.

1.2. Developments Relevant to Registration

The number of works addressing registration with learning techniques has increased in recent years. To identify them, strategic searches were performed in Scopus, Web of Science (WoS), and ArXiv. The results obtained from Scopus and WoS come from indexed journals, which means that they have passed a peer-review process. However, most of the recent works in this review were reached through ArXiv, which is a preprint repository without peer review. This is a double-edged sword. ArXiv allows disseminating the work immediately while it is being reviewed for publication in a journal. However, as some works here studied were not yet reviewed in another place, the authors had to perform a more in-depth evaluation.

Figure 2 shows the research papers published in the intersection of deep learning and registration involving 3D data over the last years in each repository. It is noticeable how the number of publications has increased over the last three years. Several search strings were employed to identify the methods surveyed in this paper. The keywords grouped by the concept are:

- To include learning-based methods: *deep learning*; *machine learning*.
- To indicate the data type employed: *3D*; *point cloud*; *mesh*.
- To gather registration proposals: *registration*; *alignment*; *transformation*; *reconstruction*.

The search strings were designed by combining the words from the previous groups by choosing one from each.

In the following sections, an analysis of learning-based approaches for registration is performed using a workflow extracted from traditional solutions. We gather in Table 1 the reviewed methods showing the individual properties for each one. These approaches allow more complex inputs such as conceptual models as well as 3D datasets. However, since each proposal uses different datasets, a fair quantitative comparison cannot be done.

Table 1. Summary of the reviewed methods. It shows the application, inputs, and outputs; the employed datasets; and the architecture of the network. The right columns record the stage of the traditional registration pipeline that the network addresses (**Ta**, **Ta**rget; **Fe**, **Fe**atures; **Ma**, **Ma**tching; **Tr**, **Tr**ansform).

			Network Details				Pipeline			
Proposal	**Year**	**Application**	**Inputs**	**Outputs**	**Datasets**	**Architecture** [1]	**Ta**	**Fe**	**Ma**	**Tr**
Yumer and Mitra [11]	2016	Shape Deformation	Point Cloud / Label	Flow / Voxel Grid	ShapeNet [12], SemEd [13]	CNN	✓	✓	✓	✓
Elbaz et al. [14]	2017	Descriptor	Depth Map	Reconstructed Depth Map	Challenging Datasets for Point Cloud Registration Algorithms [15]	AE	✗	✓	✗	✗
Li and Fan [16]	2017	MR Image Registration	(MR) Voxel Grid (x2)	Registered Voxel Grid	Alzheimer's Disease Neuroimaging Initiative (ADNI) [2]	FCN	✗	✓	✓	✓
Wang and Fang [17]	2017	3D Reconstruction from 2D Image	2D Image	3D Model (Voxel Grid)	ShapeNet [12], PASCAL3D [18], SHREC 13 [19]	GAN	✓	✓	✓	✓
Zeng et al. [20]	2017	Geometric Descriptor	Voxel Grid	Feature Vector	Analysis-by-Synthesis [21], 7-Scenes [22], SUN3D [23], RGB-D Scenes v2 [24], Halber and Funkhouser [25]	CNN	✗	✓	✗	✗
Ding and Feng [26]	2018	Multiple Point Clouds Registration (Localization)	Point Clouds	Discrete Occupancy Map	Active Vision Dataset [27]	CNN	✓	✓	✓	✓
Groueix et al. [28]	2018	Matching Deformable Shapes	Point cloud	Point Cloud	SMPL [29], SURREAL [30], SMAL [31], FAUST [32], TOSCA [33], SCAPE [34]	SDN	✗	✓	✓	✓
Gundogdu et al. [35]	2018	Garment PBS	Point Cloud / Mesh	Translation Vector	GarNet [3], SMPL [29]	PointNet	✗	✓	✓	✓
Hanocka et al. [36]	2018	Shape Alignment	Shapes (x2)	Transformed shape	ShapeNet [12], COSEG [37]	CNN	✗	✓	✓	✗
Hermoza and Sipiran [38]	2018	3D Reconstruction of Incomplete Objects	Voxel Grid (incomplete shape) / Label	Voxel Grid	ModelNet10 [39], 3D Pottery dataset [40], Custom Data	GAN	✓	✓	✓	✓
Yew and Lee [41]	2018	Descriptor	Point Clouds	Local Descriptors	Oxford RobotCar [42], KITTI Dataset [43], ETH Dataset [15]	Siam. CNN	✗	✓	✓	✗
Kuang and Schmah [44]	2018	3D Medical Image Registration	Voxel Grid (x2)	Voxel Grid	MindBoggle101 [45]	STN	✗	✓	✓	✓
Lin et al. [46]	2018	Image Compositing	RGBA Foreground / RGB Background	8 Dimensional Warp Parameter	CelebA [47], SUNCG [48]	ST-GAN	✗	✓	✓	✓
Litany et al. [49]	2018	Body Shape Completion	Partial Mesh	Completed Mesh	DFAUST [50]	VAE	✓	✓	✓	✓
Liu et al. [51]	2018	Point Cloud Flow Estimation	Point cloud (x2)	Scene flow (point level)	FlyingThings3D [52]	CNN	✗	✓	✓	✗
Mahapatra et al. [53]	2018	Multimodal Image Registration	2D medical multimodal images (x2)	Transformed Image	Retinal Images [54], Sunybrook [55]	GAN	✗	✓	✓	✓
Ofir et al. [56]	2018	Multi-spectral 2D Descriptor	RGB / NIR	Pair of Features	CIFAR-10 [57], Brown and Susstrunk [58]	Siam. CNN	✗	✓	✓	✗
Yan et al. [59]	2018	MR and TRUS Registration	MR Images (x2)	Transformation / Quality Check	Custom Data	GAN	✗	✓	✓	✓
Wang et al. [60]	2018	Force Simulation	Voxel Grid	Deformed 3D Model	Custom Data	VAE + AT	✓	✓	✓	✓
Aoki et al. [61]	2019	Point Cloud Registration	Point Clouds (x2)	Transformation	ModelNet40 [39]	MLP + PointNet	✗	✓	✓	✓
Chang and Pham [62]	2019	Point Cloud Rigid Registration	Features	Transformation	Custom Data	CNN	✗	✗	✓	✓
Guan et al. [63]	2019	Vascular Image Registration	3D CT / 2D DSA	3D Transformation (translation and rotation)	Custom Data	MCNN	✗	✓	✓	✓
Jack et al. [64]	2019	3D Reconstruction from Single Image	2D Image / Mesh	Mesh	ImageNet [65], ShapeNet [12]	CNN	✗	✓	✓	✓
Schaffert et al. [66]	2019	Correspondence Weighting	Local Features	Weights Vector	Custom Data	CNN	✗	✗	✓	✗
Smirnov et al. [67]	2019	3D Reconstruction from 2D Sketch	2D Shape	Mesh	ShapeNet [12]	CNN + MLPs	✓	✓	✓	✓
Yang et al. [68]	2019	Point Cloud Generation	Point Cloud	Point Cloud	ShapeNet [12]	AE	✓	✓	✓	✓
Wang and Solomon [69]	2019	Rigid Registration	Point Clouds (x2)	Transformation	ModelNet40 [39]	CNN	✗	✓	✓	✓
Wang et al. [70]	2019	Deformation	3D Mesh / 2D Image or Point Cloud	Mesh	ShapeNet [12]	PointNet + MLPs	✗	✓	✓	✓
Wang and Fang [71]	2019	Non-rigid Registration	Point Set (2D or 3D) (x2)	Aligned Point Set	Custom Data	MLPs	✗	✓	✓	✓
Pais et al. [72]	2020	3D Scan Registration	3D Correspondences Vector	Weights Vector / Rotation and Translation	ICL-NUIM [73], SUN3D [74]	PointNet + ResNet	✗	✗	✓	✓
Li et al. [75]	2020	Rigid Registration	Point Clouds (x2)	Transformation	ModelNet40 [39]	PointNetLK	✗	✓	✓	✓
Yuan et al. [76]	2020	Rigid Registration	Point Clouds	GMM Correspondences	ModelNet40 [39], Augmented ICL-NUIM [73,77]	PointNet	✗	✓	✓	✗
Zhang et al. [78]	2020	Multi-modal Deformable Registration	Voxelgrid (x2)	Transformation	Brain Tumor Segmentation (BraTS) [79]	GAN	✗	✓	✓	✓

[1] The architectures of some proposals are variants of the family identified in this column; [2] Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu); [3] GarNet dataset https://cvlab.epfl.ch/research/garment-simulation/garnet/.

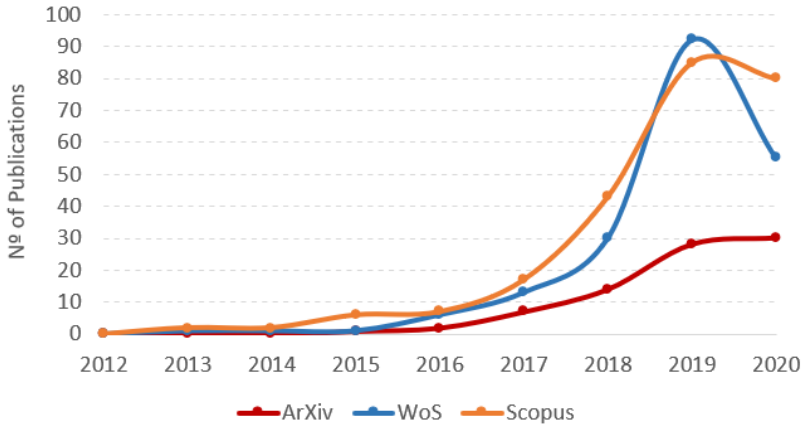

Figure 2. Research works published by year addressing registration of 3D data with learning approaches in each source (data obtained through the search engines of each source through September 2020).

2. Registration Framework

Registration aims to compute the alignment between datasets. Given two inputs $P = \{\omega_1, \omega_2, \omega_3, ..., \omega_n\}$ and $Q = \{\varphi_1, \varphi_2, \varphi_3, ..., \varphi_m\}$, a registration process finds a transformation function χ that minimizes the alignment error between P and Q, through checking the distance error between a pair of correspondences (φ_i, ω_j) of each input, as shown in Equation (1). There are different error and distance functions ($dist$), e.g., perpendicular distance rather than Euclidean distance or Huber distance, L_1 error, etc.

$$E_P = \sum_{i,j}^{n,m} gate(\varphi_i, \omega_j) * dist(\varphi_i - \chi(a, \omega_j)) \quad (1)$$

, Let χ transform each element of P, according to the transformation parameters a, with the goal of minimizing the error E_P between P and Q, while $gate(\varphi_i, \omega_j) = 1$ if the features correspond and 0 otherwise.

In the literature, the terms registration and reconstruction (or shape completion) are sometimes employed to refer to the same process. The main difference is that reconstruction is at a higher level than registration since the registration process is a part of reconstruction methods, but a reconstruction method may not perform any registration. That is, registration aims to find the transformation to align data, while the main goal of reconstruction is to obtain a virtual representation from the scene. To this end, a reconstruction method may not include a registration process, for example, those that only use a single view to obtain the virtual representation. Although these definitions previously had clear differentiation, now learning-based approaches have blurred the line between them, e.g., in [49], an algorithm designed to perform shape completion is tested by performing registration tasks.

According to Tam et al. [8], the registration process can be divided into three core components: target selection, correspondences and constraints, and optimization. This sequence has been used often by registration algorithms to find the alignment of 3D datasets. These stages are shown in Figure 1. A more detailed classification was presented by Saval-Calvo et al. [80], including pre-processing and post-processing phases:

- Pre-processing. This stage adapts the input data to meet the requirements of the algorithms.
- Target Selection. It is often necessary to differentiate between the dataset that will remain fixed and the one that will be moved towards the fixed set to perform the alignment. In the literature, different nomenclatures could be found for these fixed/moving terms such as model/data, anchor/moving, or target/source.

- Feature Extraction. This stage refers to the process of finding those landmarks or salient features that will be used to calculate the matches between sets.
- Feature Matching. It refers to the identification of corresponding features between the target and each moving data. The pair composed of pairs of features is called a correspondence.
- Pose Optimization. Here, the algorithm computes the transformation that minimizes a distance between the correspondences, aligning the sets into a common reference space.
- Post-processing. This step is highly dependant on the problem itself. It could include global optimization, such as loop-closure, data-cleaning in solid mesh estimation, surface extraction, or outlier removal.

3. Deep Learning in the Context of Registration

Deep Learning is the subfield of Machine Learning that studies Deep Neural Networks (DNN), which increases the number of hidden layers in a Neural Network (and potential layer-to-layer transformations) and computes multiple levels of abstraction. It transforms the data in a non-linear fashion by learning complex functions and transformations [81]. An extended review of the history of deep learning and its approaches can be found in [82].

In a similar way to humans, neural networks are able to *understand* the input data by extracting an abstract understanding of it [81]. Bench-Capon [83] considered the *representation of knowledge* remaining in a learning system after the training process with the following definition: *a set of syntactic and semantic conventions that makes it possible to describe things*. This representation of knowledge could be understood as a conceptual model of the object. Norman [84] defined conceptual models as an accurate, consistent, and complete representations of knowledge, coherent with the real world and physics rules. There is a gap between an observed phenomenon and the mathematical model. According to Nersessian [85], mental models are located in this gap, but they can be incomplete or unscientific. A mathematical model is also a conceptual model, which is an external representation that facilitates the comprehension of a teaching system. It is functional and coherent with scientific knowledge [86].

This conceptual knowledge can improve alignment problems. Traditional registration approaches have different challenges, that lead into one general limitation: the lack of generalization of these algorithms. Usually they are highly dependent on the correspondences between the input datasets. With the development of DL, the *remaining knowledge*, defined before as a set of syntactic and semantic conventions, could be considered as a conceptual model, that, in the case of registration processes, could be a target to align with spatial data. Theoretically, the idea of the conceptual model allows to differentiate the input data of a registration process into defined or non-defined models. The defined aspects are models that represent specific spatial data (commonly 2D or 3D) while a non-defined model is a generalization of a dataset produced by a learning system, e.g., the concept of a ball, those properties that make an object a ball, rather than the specific instance of a ball itself using geometrical aspects.

The conceptual models have also been applied in registration, for example, in the work of Yumer and Mitra [11], in which the network learns properties of objects, being able to know what a more sporty car looks like or a more comfortable chair is, and modifying a 3D model to fit those properties while preserving the main features of the original data. With this approach, three combinations of input information are possible: *defined model/defined model*, *defined model/conceptual model*, and *conceptual model/conceptual model*. This taxonomy is shown in Figure 3. The classical algorithms for registration are included in the first of the possibilities, one input is used as a target or as a reference set whilst the other is transformed to be aligned with the first, but always with defined data. By contrast, the use of neural networks for registration results in other combinations where conceptual models are included. Those models need not be specifically defined, e.g., they can be synthesized by a trained network with the learned features coming from the training data. Then, these features can be used

in the registration process afterward or even in the same network. In any case, there is no need to understand the working space of the network. Its internal representation is alien to human knowledge.

The combination of two conceptual models could be possible with the growth of *Imagination Machines* proposed by Mahadevan [87], which aims to provide artificial intelligence systems flexibility and connections between the learned aspects through training processes not based on labels and classifications of the input data.

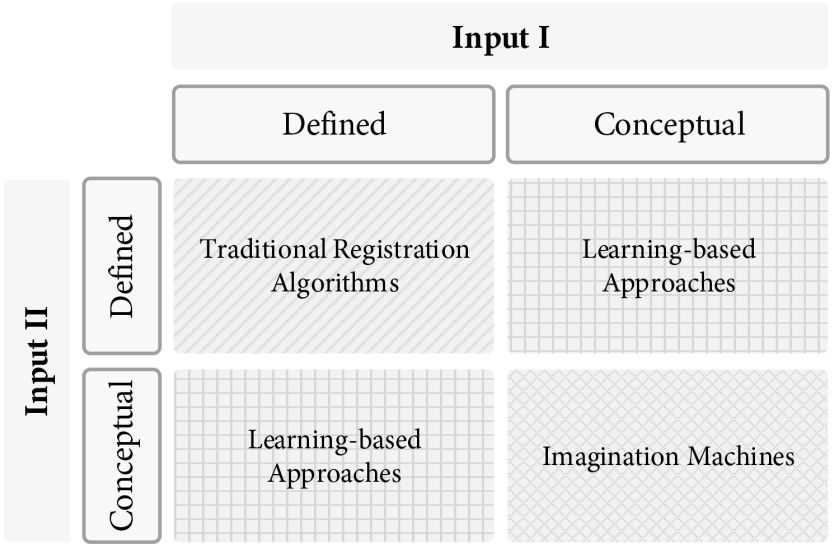

Figure 3. The taxonomy present in registration algorithms as a result of the intersection between defined and conceptual/non-defined features.

4. Review of Learning-Based Approaches for Registration

To show the advancements of neural networks applied to registration problems, an analysis of recent works in the intersection of these fields is performed. To establish a comparative framework between these new approaches, we use the workflow abstracted from traditional solutions, as introduced in Section 1. This workflow is divided into four stages: target, features, matching, and optimization. In this section, an analysis of learning approaches that use some or all of these stages of the workflow is performed.

The reviewed methods are shown in Table 1, classified regarding the traditional phases of a registration workflow. The first columns are the application, inputs, outputs, employed datasets, and architecture of each method. The final columns indicate the parts of the traditional registration process implemented in the DN. In this way, the target column refers to the need for the method to have anchor data as a target to perform the alignment. If there is no checkmark in a column, it means the method requires a defined target as input, otherwise the target is a conceptual model inside the network. This is possible if the generalization of the training data is implicit in the network knowledge, i.e., the main properties from the inputs are learned by the network. For example, it is possible to think of a specific instance of a chair or think of a chair as a concept, with those properties that a chair must fulfill to be considered as such.

The feature column indicates the ability of the method to find features in the data using a neural network, such as the work of Ofir et al. [56]. The next step in the workflow is the matching between features. There are some proposals which train a network to be able to check the accuracy of the correspondences; Pais et al. [72] proposed a network to align two datasets given the features and the matching between them. The network determines if the features are correct, removing some of them if necessary. The last column indicates the ability of the learning approach managing the geometric operations that align datasets, such as computing the camera pose [26] or the transformation parameters [35].

In the recent state-of-the-art methods, we find methods that carry out the whole registration, i.e., they cover the main parts of the traditional pipeline of registration, as well as other methods that implement only some stages of the pipeline. This classification of the analyzed works has been done as a way to compare them in a common framework using a traditional perspective of the registration methods, going into detail in the key aspects of each method according to the stage where it contributes a novel method.

4.1. Target Level

At the target level, there are methods that generalize from the training process, exploring the idea of the conceptual model. This enables registration using learning approaches. For instance, the work of Yumer and Mitra [11] uses semantically deforming shapes in 3D through free-form deformation using lattices. It does not use a target model to deform a mesh, the properties of the target are learned by the proposed network. The network is able to perform non-rigid deformations over 3D models to fit a given semantic property. As a result, it provides the deformed 3D model as well as the deformation flow of the data to fit that model while preserving original details. Similarly, a key aspect of the unsupervised learning approach proposed by Ding and Feng [26] is that there is no target for the registration process. The network is capable of locating each input point cloud in a global space, solving SLAM problems in which multiple point clouds have to be registered rigidly. The employed architecture is discussed in the following sections.

Adversarial training (AT), Autoencoders (AE), and Generative Adversarial Networks (GANs) are used in some works for extracting conceptual descriptions. For instance, Wang and Fang [17] employed an adversarial approach with CNN networks to reconstruct a 3D model of an object from its 2D image. The key aspect of this work is the combination of a 2D autoencoder-based network with a deconvolutional network. The first network transforms the input image to the latent space, while the second transforms from the latent space to 3D space, acting as a 3D generator. It is an unsupervised generative neural network that accurately predicts 3D volumetric objects from single real-world 2D images. The network has learned multiple objects and internally performs the registration between the image and the conceptual model. In a similar way, Hermoza and Sipiran [38] also used a GAN network for predicting the missing geometry of damaged archaeological objects, indicating the reconstructed object in a voxel grid format and a label designating its class. Its network architecture combines a completion loss and an improved Wasserstein GAN loss.

Smirnov et al. [67] proposed a method to generate a 3D model from a 2D sketch. The 3D models are defined by a set of parametric patches. They employed an encoder-style architecture using convolutional layers and residual blocks that generates a series of 3D patches from the sketch, then a set of MLPs carry out the intersection between patches. In this work, registration between different spaces is performed with the provided sketch and the internal knowledge that comes from the training procedure. Similarly, the generative model of Yang et al. [68] uses a variation of an autoencoder architecture to generate 3D point clouds by modeling them as a distribution of distributions. Concretely, their method learns the distribution of shapes at the first level, and the distribution of points given a shape at the second level. As a result, the method is able to generate points as a given shape by parameterizing the transformation of points from an initial Gaussian distribution of them. Moreover, Variational Autoencoders (VAEs) are being used in an adversarial training framework, such as the work of Wang et al. [60]. In this case, they are employed for predicting structural deformations produced by forces given a single depth image and the conditions of the input, which includes properties of the material, the strength of the force, its location, etc. The generator predicts the force over a 3D model, and the discriminator, used for training, should determine if the applied force comes from the generator or from the ground-truth. This approach enables the network to learn non-rigid deformations and it can generalize the deformations to unknown objects having into account properties of the materials. Other approaches are able to train a variational autoencoder with graph convolutional operations for completing missing data from partial body shapes while dealing with non-rigid deformations

[49]. They are able to identify the output space of the generator that best aligns with the partial input. Partial shapes are completed by deforming a randomly generated shape to be aligned with a partial input. This approach is robust to non-rigid deformations and has the ability to reconstruct missing data. It shows topology understanding by the encoder–decoder architecture.

4.2. Feature and Matching Level

Learning approaches have demonstrated successful results in performing feature extraction and matching for registration purposes. Auto-encoders have been used for feature extraction. For instance, Elbaz et al. [14] used in their proposal for point cloud registration a Deep Auto Encoder (DAE) for extracting low-dimensional descriptors from large scale point clouds. The training of the DAE is unsupervised, and it is able to extract a compact representation from depth maps that capture the significant geometric properties of the input data. Yuan et al. [76] proposed the method DeepGMR to perform the registration by matching points to a probability distribution whose parameters are estimated by a neural network from the input point clouds. That network learns latent correspondences between points and Gaussian Mixture Model (GMM) components that are pose-invariant. The network estimates point-to-component correspondences, following two compute blocks to obtain the GMM parameters and the transformation. Groueix et al. [28] introduced Shape Deformation Networks (SDNs) in an encoder–decoder architecture for matching deformable shapes, where the encoder is able to extract a global shape descriptor from a 3D model, while the decoder can transform the extracted descriptor into another model. The SDN is able to learn to deform a template shape to be aligned to targets with the articulated restriction. Concretely, the encoder SDN learns the deformation parameters and degrees of freedom to deform the template. This work shows that an encoder–decoder architecture to generate human shape correspondences can compete with state-of-the-art methods.

Convolutional Neuronal Networks have also been used for feature extraction. Hanocka et al. [36] propose ALIGNet, an unsupervised network to align either 2D or 3D shapes with an incomplete target. The network learns to extract the features to match both shapes and compute Free Form Deformation (FFD) grids. It is trained with a shape alignment loss by comparing the overlap between the source and the target for learning the FFD parameters. Ofir et al. [56] developed a learning-based method to register multi-spectral images (visible and Near-Infra-Red images). They employed a learning approach for extracting features of both images and matching them. For that purpose, their proposal is based on an asymmetric (different weights) Siamese Convolutional Neural Network, one for each spectral channel. The networks minimize the Euclidean distance between the two descriptors. With a similar network architecture, Yew and Lee [41] proposed 3DFet-Net, which finds features and descriptors as well as correspondences for later registration. They used coarsely annotated point clouds with GPS/INS absolute pose. It is based on a three-branch Siamese architecture that uses PointNet++ [88]. Each branch takes an entire point cloud as input. The network is trained with a set of triplets containing the anchor and positive and negative point clouds. Positive point clouds are those with a distance to the anchor below a threshold, and negative point clouds are far away from the anchor. Each branch has a detector and descriptor network. Both networks for each branch share the same inputs. The detector network predicts an orientation and an attention weight for each branch. Then, the descriptor network rotates the input to a canonical configuration and computes the features that will be aligned with the other branch through a triplet loss. That loss aims to minimize the difference between the anchor and positive point cloud and maximize the difference between the anchor and the negative point cloud.

To address the problem of inaccurate correspondences, Schaffert et al. [66] employed a modified PointNet [89] architecture for weighting individual correspondences in a 2D/3D rigid registration process on X-ray images. They employed a modified PointNet to process points individually to obtain global information. The authors included a second MLP which processes correspondences containing global and local information. This modified network is able to weight individual correspondences based on their geometrical properties and similarity as well as global properties.

Chang and Pham [62] presented a 3D point set registration framework with two stages to cover the problem of coarse-to-fine registration. Two descriptors are proposed, one for rough and one for fine orientation extraction, SSPD and 8CBCP, respectively. SSPD that is a normalized voxel grid and 8CBCP describes the orientation using an 8dx3 matrix obtained from the data in the eight sub-quadrants of the bounding box around the 3D points. They used two consecutive CNNs using those descriptors. The first CNN receives as input the SSPD descriptor and is meant to estimate the coarse alignment. Then, the 8CBCP descriptor is computed over the output and introduced in a second CCN that performs a more accurate alignment. In this proposal, the CNNs only estimate the rotation, and the translation is afterward obtained.

Convolutions used in most of the deep learning networks operate over a neighborhood of the data; thus, structured inputs are required. 3D point clouds are unorganized datasets that are challenging to operate by convolution-based networks, a problem that led to much research on this topic. Some state-of-the-art proposals tackle this problem by voxeling the point cloud [90] but these approaches are not efficient since points are sparse and a large percentage of voxels are empty and details can be lost. Others try to extract geometric features from point clouds, e.g. Xu et al. [91] used so-called SpiderConv filters that are parameterized functions of specific radius applied over the point cloud. The ELF-Nets of Lee et al. [92] proposed the Extended Laplacian Filter that is a combination of a two-state filter, one for the center point and one for neighbors, with a scalar weighting function that represents the relative importance of the points. This last approach uses fewer parameters than SpiderConvs. For managing 3D data, Zeng et al. [20] employed a reduced set of voxels of TDF (truncate distance function) containing an interest point of a point cloud. The TDF is the distance from the center of the voxel to the nearest point. This is used as input of a convolutional network which extracts a 512-dimensional feature representation. The result is a geometric descriptor in which the network is able to generalize to other tasks and resolutions. However, according to Liu et al. [51], the CNN does not provide good results when working with point clouds due to their irregular structure. For this purpose, they employed a modified version of the PointNet++ [88] architecture, a network that learns hierarchical features. With that network, they propose a network architecture that learns to predict scene flow as translational motion vectors for each point. The proposed architecture has three modules: point feature learning, point mixture, and flow refinement. It includes a *flow embedding layer* that learns to aggregate geometric similarities and spatial relations of points for motion encoding. Pais et al. [72] presented a network architecture with two main blocks, the classification block fed with pairs of corresponding 3D points and giving as a result features for each correspondence using 12 ResNets [93], which remove outlier correspondences. The registration block gets the resulting features from the previous stage and produces a six-variable output for rotation and translation obtained with a context normalization layer along with convolutional one and two fully connected layers. This method works on point correspondences. It is efficient and outperforms traditional approaches.

4.3. Transformation Level

Some works have been successfully employed neural networks for learning and applying geometric transformations. Some of these achievements have been done using GAN architectures or variants of them. For instance, Lin et al. [46] demonstrated good results using neural networks for finding realistic geometric transformations for 2D image compositing. Image compositing refers to overlap images coming from different scenes; thus, achieving a good realism implies a good transformation to minimize the appearance and geometric differences. For this purpose, they propose a GAN architecture using Spatial Transform Networks (STNs) [94], named ST-GAN. According to the authors of ST-GAN, this idea could be extended to other image alignment tasks. STNs have shown good results resolving geometric variations; thus, with this architecture, the network learns to perform realistic geometric warps, demonstrating potential 3D capabilities. Yan et al. [59] also used GANs to carry out the registration of magnetic resonance (MR) and transrectal ultrasound (TRUS) images as well as evaluate the provided result. The generator network provides the transformation parameters to align

both inputs, while the discriminator performs a quality evaluation of that alignment. Mahapatra et al. [53] used GANs for deformable multimodal medical image registration in 2D. The network outputs a transformed image and also a deformation field. Similarly, in three-dimensional space, Hermoza and Sipiran [38], as referenced above, performed the reconstruction of incomplete archaeological models also using GANs, in which the generator network provides a reconstructed model. Zhang et al. [78] also proposed a registration method based on a GAN architecture with a gradient loss which can manage local structure information across modalities. This makes it more robust against large deformations, noise, and blur.

Ding and Feng [26] managed multiple point clouds registration using DNNs. They approached this problem by including two networks, a localization network named L-Net, and an occupancy map network, M-Net. The L-Net estimates the sensor pose for a given point cloud, sharing some optimization parameters between the input point clouds. The goal of this network is to estimate the sensor pose in a global frame. To do that, the L-Net network is divided into a feature extraction module followed by an MLP that outputs the sensor pose. The feature extraction module employed depends on the input format of the point cloud. If it is an organized point cloud, a CNN is employed for that purpose. If not, the features are extracted using PointNet [89]. Later, the M-Net receives those location coordinates in the global space and retrieves the discrete occupancy map. Besides, the L-Net network locates each input point cloud in a global space; there is no target for the alignment. With a similar architecture, Wang et al. [70] presented 3DN, a combination of PointNet and MLPs that deforms 3D meshes to resemble a target, given in the form of a 2D image or point cloud, as close as possible preserving the properties of the source. The proposal extracts global features from both source and target inputs using CNN/PointNet. Next, those features are used to estimate the per-vertex displacement with an 'offset decoder'. To overcome the problem of tessellation differences, an intermediate sampled point cloud is calculated from both source and target. They employed a combination of four different loss functions, measuring the similarity between the deformed source and the target, symmetry, local geometric details, and self-intersections. This work proposes an end-to-end network architecture for mesh deformation.

Using autoencoders, Groueix et al. [28], with their SDNs introduced before, replicated the shape of a body previously encoded in a given template. For 3D medical image non-rigid registration, Kuang and Schmah [44] employed an architecture inspired by STNs extending the works of Shan et al. [95] and Balakrishnan et al. [96]. The network takes a pair of volumes and predicts the displacement fields needed to register source to target. According to the authors, it improves the results compared to U-net [97] and VoxelMorph [96]. This method produces deformations with fewer regions of non-invertibility where the surface folds over itself. To achieve this, they employed an explicit anti-folding regularization to penalize *foldings*, which are the spatial locations where the deformation is non-invertible and is indicated by a negative determinant of the Jacobian matrix.

With convolutional networks, Jack et al. [64] performed the 3D reconstruction from a single 2D image by learning to apply large deformations and compelling mesh reconstructions by inferring Free Form Deformation (FFD) parameters. They employed a lightweight CNN based on the MobileNet architecture [98] to infer FDD parameters to deform a template and infer a 3D mesh of the given image. As a result, the network learns how to deform a given template to match features present in a 2D image with finer geometry than other methods working with voxel grids and point clouds, because there is no discretization.

Guan et al. [63] proposed a multi-channel CNN (MCNN) for deformable registration of CT scans with digital subtraction angiography (DSA) images of the cardiovascular system. The network is composed of several sub-networks that converge before the fully connected layers. They named this architecture as a multi-channel convolutional neural network. They employed a CNN model based on the VGG network combined with a vascular diameter variation model to directly regress and predict transformation parameters. With this architecture, each channel of the MCNN process a different phase of the vascular deformation cycle, comparing the results of each to choose the best

result. Li and Fan [16] employed Fully Convolutional Networks (FCNs) to optimize and learn spatial transformations between pairs of images to be non-rigidly registered. Their method works with medical images at the voxel level and, according to the authors, it improves the results of STNs, which cannot manage small transformations. The spatial transformation between pairs of images is obtained directly by maximizing an image-wise similarity metric, similar to traditional approaches. The use of FCNs facilitates the voxel-to-voxel prediction of deformation fields, which also allows learning small transformations.

Gundogdu et al. [35] proposed a method for 3D garment fitting on bodies. To extract global features of the body model, they employed a PointNet [89] but with leaky ReLUs with a slope of 0.1. After that, a second stream, composed by six residual blocks, is used to extract features from the garment mesh and also take as input the previous global body features. Thirdly, the features provided by both networks are merged employing four Multi Layer Perceptron (MLP) blocks shared by all points. The final MLP block outputs a vector with the 3D translation information. With this method, the authors achieved results nearly as accurate as a Physics-Based Simulation (PBS), but less time-consuming.

Similarly, PointNetLK [61] is a method for 3D rigid registration which modifies the Lucas and Kanade (LK) [99] algorithm integrated with PointNet. The process is mainly divided into two steps: initially, two 3D point sets are passed through a shared Multi Layer Perceptron of the two inputs and a symmetric pooling function. Second, the transformation is obtained and applied to the moving point cloud. The whole procedure is iteratively repeated until a minimum quality threshold is reached. According to the authors, this method exhibits remarkable generalization of unseen objects and shape variation due to the encoding of the alignment process in the network architecture that only needs to learn the PointNet representation. Li et al. [75] proposed a deterministic PointNetLK method to improve the generalization by using analytical gradients. Wang and Fang [71] presented CPD-Net, a network architecture that performs non-rigid registration under the concept of learning a displacement vector function that estimates the geometric transformation. The pipeline is decomposed into three main components: 'Learning Shape Descriptor' with a MultiLayer Perceptron (MLP) that learns descriptors from the input source and target point sets; 'Coherent PointMorph' that is a three MLPs block fed with the two descriptors concatenated with the source data points; and the 'Point Set Alignment', where the loss function is defined to determine the quality of the alignment. Deep Closest Point [69] registers two point clouds by first embedding them into high-dimensional space using DGCNN [100] to extract features. After that, contextual information is estimated using an attention-based module that provides a dependency term between the feature sets, i.e., one set is modified in a way that is knowledgeable about the structure of the other. Finally, alignment is obtained using a differentiable Singular Value Decomposition (SVD) layer, which seems to provide better results than an MLP. This proposal also includes a "pointer generation" that provides a probabilistic approach to generate a soft map between the two features sets to minimize the problem of falling into a local minima.

Going back to the work of Pais et al. [72] cited above, the component performing the alignment uses a CNN that receives as input features extracted from the selected correspondences at different stages of the previous components composed by ResNet blocks. The registration block receives as input features previously extracted and outputs the transformation parameters.

4.4. Summary

Through this section, an overview of neural networks that perform alignment or registration tasks is provided. A perspective based on the traditionally employed pipeline in registration methods is used to analyze the proposals. From the analyzed works (summarized in Table 1), it is possible to observe that the extraction and matching of features are tasks widely explored with neural networks because they are common points with other problems such as object classification or recognition.

However, it is not common to find neural networks with the ability to manage geometric information for applying transformations on data to meet some requirements.

There are approaches dealing with some parts of the pipeline or the whole registration. Interestingly, the proposals that compute the transformation for the alignment also perform the matching of features. That means there are no proposals performing only the calculation of the transformation. In terms of deformable alignment, there exist specific proposals for learning deformations or non-rigid registration with networks, such as the SDNs, but is a current topic under research.

In addition, it is noticeable that most of the analyzed neural networks employ a GAN, a CNN, or a variation of them, but there are a few works with a GAN architecture managing 3D data. However, this point is under active study because this kind of data requires many resources. As occurs with 2D images, the main solution is to use discrete input data. Similar to pixels in 2D, voxel grids are the common solution for 3D. However, in some situations, it is important to work at point level, e.g., for estimating deformation flow. Although there are some proposals working with point clouds as input data, most of them require an organized point cloud or a limitation in the number of points.

From the reviewed methods, it is possible to extract the key innovations that are relevant to registration problems. Table 2 summarizes the key contributions of the reviewed methods classified according to the stage in which they are relevant.

Table 2. Summary of the key advantages of the reviewed methods.

target selection
• Reconstruction of 3D models from a single 2D image using encoder–decoder architectures [17,67] • Leverage partial 3D observations to generate complete 3D models (mesh completion) using Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) for predicting missing geometry [38,49] • The use of Adversarial Training (AT) to predict structural deformations on 3D meshes given a multi-modal input (e.g., forces, material properties, visual imaging) [60]
features and matching
• Registration of global and local point clouds with a generic Deep Auto Encoder (DAE) architecture regardless of the input data and its source [14] • The Shape Deformation Network (SDN) is an autoencoder architecture able to deal with deformable shapes extracting global shape descriptors [28] • Capability for multi-spectral registration using Asymmetric Siamese Convolutional Networks [56] • Filtering inaccurate features correspondences based on geometrical and global properties to minimize their influence in the registration process [66,72] • Registration process speedup using a two-staged approach based on Convolutional Neural Networks (CNNs) to solve coarse-to-fine registration problems [62]
transform
• Using Spatial Transformer Networks (STNs) to deal with geometrical variability by means of learned invariance to transformations [44,46,94] • To align multimodal inputs with transformations generated and evaluated in an adversarial fashion [59] • Ability to preserve detailed geometric information by using a CNN to infer Free Form Deformation (FFD) parameters for 3D template-image matching [64] • Unsupervised registration of multiple point clouds in a global frame of reference [26] • Outperforming single channel CNNs with a multi-channel approach for real-time deformable registration [63] • Cloth fitting by modeling 3D body-garment interaction in real time surpassing Physics-Based Simulation (PBS) [35]

5. Conclusions

In this work, approaches in the intersection between registration and deep neural networks are reviewed. It is important to remark that a large part of the works was reached through the ArXiv repository, which at the moment, even though is not peer-reviewed, leads the advancements in many fields such as artificial intelligence and deep learning. For this reason, extra effort was required to select the papers. However, this approach enables including the most recent works in the scope of this paper.

Registration aims to calculate a common reference for two or more datasets. This field has been widely studied, but recent techniques in machine learning are being combined with registration algorithms to increase the capabilities of the proposals. These techniques include neural networks with many hidden layers, also known as deep learning, and its novelty remains in the ability to learn representations from huge amounts of data at different levels of abstraction. Applied to registration, this paradigm allows managing higher-level understanding problems that are more related to conceptual knowledge of the scene rather than the geometric properties. We name this paradigm Deep Registration Networks (DRNs) to identify the branch of artificial intelligence exploring solutions for alignment problems using DNNs.

The contribution of this work is a review of registration methods based on deep networks. To do that, the learning approaches for registration are reviewed and classified using a novel framework extracted from the traditional registration pipeline. The review clearly identifies the current efforts and existing gaps in the intersection between registration and learning algorithms. Moreover, the positive influence of the internal representations modeled by learning approaches on the registration process is clearly identified.

As a result, an overall view of this new subfield is provided, setting out different architectures and solutions that are being provided by the authors. A summary of the different methods is shown in Table 1 with the inputs, outputs, architectures, and datasets employed to address the problems. Besides, an analysis of each method has been made using a traditional perspective of the pipeline employed in registration algorithms. As the main contribution of this work, we provide a framework to understand the learning methods for registration. In these new methods, the stages of the pipeline are not so clearly defined as they are in a traditional approach, because some processes are computed directly and implicitly by the network, e.g., the extraction and matching of features. However, an advantage of the learning approaches is that they are suitable for real-time problems. The higher computational needs of a neural network are at the training phase, which is performed once. After that, the data processing is relatively fast for real-time applications.

From our perspective, it is clear that researchers are still exploring different paradigms, and no single approach is so far the preferred one. Whether the learning-based approaches will enable significant improvements over traditional registration approaches is still an open question. To help assess whether convergence in the literature is happening, we analyzed the approaches using k-means and SOM networks to find clusters of methods sharing characteristics. However, no significant clusters were found, suggesting that convergence has not yet happened. The metrics employed to evaluate each method are different, some of them are even ad hoc solutions. In addition, there is a lack of standard benchmarks as well as common datasets to compare/evaluate the methods. For this reason, a comparison between methods is not a contribution as it would not allow extracting relevant conclusions.

To conclude, we find that most current approaches can be analyzed using concepts from the four stages of registration identified in Figure 1, which enable the recognition, registration, and reconstruction of objects. Although the four stages are evident in the traditional algorithms, with the rise of deep learning, we believe that it will be possible to deal with more complex registration problems, e.g., at the conceptual level.

Funding: This work was supported by the Spanish State Research Agency (AEI) and the European Regional Development Fund (FEDER) under project TIN2017-89069-R. This work was also supported by a Valencian Regional project (GV/2020/056), two Valencian Grants for Ph.D. studies (ACIF/2017/223 and ACIF/2018/197), and two Valencian Grants for predoctoral internships (BEFPI/2020/001 and BEFPI/2020/068).

References

1. Villena-Martinez, V.; Fuster-Guillo, A.; Saval-Calvo, M.; Azorin-Lopez, J. 3D Body Registration from RGB-D Data with Unconstrained Movements and Single Sensor. In *International Work-Conference on Artificial Neural Networks*; Springer International Publishing: Cham, Switzerland, 2017; pp. 317–329. doi:10.1007/978-3-319-59147-6_28.
2. Zeman, R.K.; Davros, W.J.; Berman, P.; Weltman, D.I.; Silverman, P.M.; Cooper, C.; Evans, S.R.; Buras, R.R.; Stahl, T.J.; Nauta, R.J. Three-dimensional models of the abdominal vasculature based on helical CT: usefulness in patients with pancreatic neoplasms. *Am. J. Roentgenol.* **1994**, *162*, 1425–1429. doi:10.2214/ajr.162.6.8192012.
3. Boldea, V.; Sharp, G.C.; Jiang, S.B.; Sarrut, D. 4D-CT lung motion estimation with deformable registration: Quantification of motion nonlinearity and hysteresis. *Med. Phys.* **2008**, *35*, 1008–1018. doi:10.1118/1.2839103.
4. Cuevas-Velasquez, H.; Li, N.; Tylecek, R.; Saval-Calvo, M.; Fisher, R.B. Hybrid Multi-camera Visual Servoing to Moving Target. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018. doi:10.1109/iros.2018.8593652.
5. Zhao, Z.Q.; Zheng, P.; Xu, S.T.; Wu, X. Object Detection With Deep Learning: A Review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. doi:10.1109/tnnls.2018.2876865.
6. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Martinez-Gonzalez, P.; Garcia-Rodriguez, J. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* **2018**, *70*, 41–65. doi:10.1016/j.asoc.2018.05.018.
7. Oprea, S.; Martinez-Gonzalez, P.; Garcia-Garcia, A.; Castro-Vargas, J.A.; Orts-Escolano, S.; Garcia-Rodriguez, J.; Argyros, A. A Review on Deep Learning Techniques for Video Prediction. *arXiv* **2020**, arXiv:2004.05214.
8. Tam, G.K.L.; Cheng, Z.Q.; Lai, Y.K.; Langbein, F.C.; Liu, Y.; Marshall, D.; Martin, R.R.; Sun, X.F.; Rosin, P.L. Registration of 3D Point Clouds and Meshes: A Survey from Rigid to Nonrigid. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 1199–1217. doi:10.1109/tvcg.2012.310.
9. Zhu, H.; Guo, B.; Zou, K.; Li, Y.; Yuen, K.V.; Mihaylova, L.; Leung, H. A Review of Point Set Registration:From Pairwise Registration to Groupwise Registration. *Sensors* **2019**, *19*, 1191. doi:10.3390/s19051191.
10. Salvi, J.; Matabosch, C.; Fofi, D.; Forest, J. A review of recent range image registration methods with accuracy evaluation. *Image Vis. Comput.* **2007**, *25*, 578–596. doi:10.1016/j.imavis.2006.05.012.
11. Yumer, M.E.; Mitra, N.J. Learning Semantic Deformation Flows with 3D Convolutional Networks. In *ECCV*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 294–311. doi:10.1007/978-3-319-46466-4_18.
12. Chang, A.X.; Funkhouser, T.; Guibas, L.; Hanrahan, P.; Huang, Q.; Li, Z.; Savarese, S.; Savva, M.; Song, S.; Su, H.; et al. Shapenet: An information-rich 3d model repository. *arXiv* **2015**, arXiv:1512.03012.
13. Yumer, M.E.; Chaudhuri, S.; Hodgins, J.K.; Kara, L.B. Semantic shape editing using deformation handles. *ACM Trans. Graph.* **2015**, *34*, 1–12. doi:10.1145/2766908.
14. Elbaz, G.; Avraham, T.; Fischer, A. 3D Point Cloud Registration for Localization Using a Deep Neural Network Auto-Encoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.265.
15. Pomerleau, F.; Liu, M.; Colas, F.; Siegwart, R. Challenging data sets for point cloud registration algorithms. *Int. J. Rob. Res.* **2012**, *31*, 1705–1711. doi:10.1177/0278364912458814.
16. Li, H.; Fan, Y. Non-rigid image registration using fully convolutional networks with deep self-supervision. *arXiv* **2017**, arXiv:1709.00799.
17. Wang, L.; Fang, Y. Unsupervised 3D reconstruction from a single image via adversarial learning. *arXiv* **2017**, arXiv:1711.09312.
18. Xiang, Y.; Mottaghi, R.; Savarese, S. Beyond PASCAL: A benchmark for 3D object detection in the wild. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO, USA, 24–26 March 2014. doi:10.1109/wacv.2014.6836101.

19. Li, B.; Lu, Y.; Godil, A.; Schreck, T.; Aono, M.; Johan, H.; Saavedra, J.M.; Tashiro, S. SHREC'13 Track: Large Scale Sketch-Based 3D Shape Retrieval. *EG 3DOR*; Eurographics: Aire-la-Ville, Switzerland, 2013, pp. 89–96. doi:10.2312/3DOR/3DOR13/089-096.
20. Zeng, A.; Song, S.; NieBner, M.; Fisher, M.; Xiao, J.; Funkhouser, T. 3DMatch: Learning Local Geometric Descriptors from RGB-D Reconstructions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.29.
21. Valentin, J.; Dai, A.; Niessner, M.; Kohli, P.; Torr, P.; Izadi, S.; Keskin, C. Learning to Navigate the Energy Landscape. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016. doi:10.1109/3dv.2016.41.
22. Shotton, J.; Glocker, B.; Zach, C.; Izadi, S.; Criminisi, A.; Fitzgibbon, A. Scene Coordinate Regression Forests for Camera Relocalization in RGB-D Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013. doi:10.1109/cvpr.2013.377.
23. Xiao, J.; Owens, A.; Torralba, A. SUN3D: A Database of Big Spaces Reconstructed Using SfM and Object Labels. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 1–8 December 2013. doi:10.1109/iccv.2013.458.
24. Lai, K.; Bo, L.; Fox, D. Unsupervised feature learning for 3D scene labeling. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014. doi:10.1109/icra.2014.6907298.
25. Halber, M.; Funkhouser, T. Fine-to-Coarse Global Registration of RGB-D Scans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.705.
26. Ding, L.; Feng, C. DeepMapping: Unsupervised Map Estimation From Multiple Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. doi:10.1109/cvpr.2019.00885.
27. Ammirato, P.; Poirson, P.; Park, E.; Kosecka, J.; Berg, A.C. A dataset for developing and benchmarking active vision. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017. doi:10.1109/icra.2017.7989164.
28. Groueix, T.; Fisher, M.; Kim, V.G.; Russell, B.C.; Aubry, M. 3D-CODED: 3D Correspondences by Deep Deformation. In *ECCV*; Springer International Publishing: Cham, Switzerland, 2018; pp. 235–251. doi:10.1007/978-3-030-01216-8_15.
29. Loper, M.; Mahmood, N.; Romero, J.; Pons-Moll, G.; Black, M.J. SMPL. *ACM Trans. Graph.* **2015**, *34*, 1–16. doi:10.1145/2816795.2818013.
30. Varol, G.; Romero, J.; Martin, X.; Mahmood, N.; Black, M.J.; Laptev, I.; Schmid, C. Learning from Synthetic Humans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.492.
31. Zuffi, S.; Kanazawa, A.; Jacobs, D.W.; Black, M.J. 3D Menagerie: Modeling the 3D Shape and Pose of Animals. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.586.
32. Bogo, F.; Romero, J.; Loper, M.; Black, M.J. FAUST: Dataset and Evaluation for 3D Mesh Registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 24-27 June 2014, Columbus, OH, USA. doi:10.1109/cvpr.2014.491.
33. Bronstein, A.M.; Bronstein, M.M.; Kimmel, R. *Numerical Geometry of Non-Rigid Shapes*; Springer Science & Business Media: New York, NY, USA, 2008.
34. Anguelov, D.; Srinivasan, P.; Koller, D.; Thrun, S.; Rodgers, J.; Davis, J. SCAPE. *ACM SIGGRAPH 2005*; ACM Press: New York, NY, USA, 2005. doi:10.1145/1186822.1073207.
35. Gundogdu, E.; Constantin, V.; Seifoddini, A.; Dang, M.; Salzmann, M.; Fua, P. Garnet: A two-stream network for fast and accurate 3d cloth draping. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 8739–8748.
36. Hanocka, R.; Fish, N.; Wang, Z.; Giryes, R.; Fleishman, S.; Cohen-Or, D. ALIGNet. *ACM Trans. Graph.* **2018**, *38*, 1–14. doi:10.1145/3267347.
37. Wang, Y.; Asafi, S.; van Kaick, O.; Zhang, H.; Cohen-Or, D.; Chen, B. Active co-analysis of a set of shapes. *ACM Trans. Graph.* **2012**, *31*, 1. doi:10.1145/2366145.2366184.

38. Hermoza, R.; Sipiran, I. 3D Reconstruction of Incomplete Archaeological Objects Using a Generative Adversarial Network. In Proceedings of the Computer Graphics International 2018, Bintan Island, Indonesia, 11–14 June 2018. doi:10.1145/3208159.3208173.
39. Wu, Z.; Song, S.; Khosla, A.; Yu, F.; Zhang, L.; Tang, X.; Xiao, J. 3D ShapeNets: A deep representation for volumetric shapes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015. doi:10.1109/cvpr.2015.7298801.
40. Koutsoudis, A.; Pavlidis, G.; Liami, V.; Tsiafakis, D.; Chamzas, C. 3D Pottery content-based retrieval based on pose normalisation and segmentation. *J. Cult. Herit.* **2010**, *11*, 329–338. doi:10.1016/j.culher.2010.02.002.
41. Yew, Z.J.; Lee, G.H. 3DFeat-Net: Weakly Supervised Local 3D Features for Point Cloud Registration. In *ECCV*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 630–646. doi:10.1007/978-3-030-01267-0_37.
42. Maddern, W.; Pascoe, G.; Linegar, C.; Newman, P. 1 year, 1000 km: The Oxford RobotCar dataset. *Int. J. Rob. Res.* **2016**, *36*, 3–15. doi:10.1177/0278364916679498.
43. Geiger, A.; Lenz, P.; Urtasun, R. Are we ready for autonomous driving? The KITTI vision benchmark suite. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012. doi:10.1109/cvpr.2012.6248074.
44. Kuang, D.; Schmah, T. FAIM – A ConvNet Method for Unsupervised 3D Medical Image Registration. In *MLMI*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 646–654. doi:10.1007/978-3-030-32692-0_74.
45. Klein, A.; Tourville, J. 101 Labeled Brain Images and a Consistent Human Cortical Labeling Protocol. *Front. Neurosci.* **2012**, *6*. doi:10.3389/fnins.2012.00171.
46. Lin, C.H.; Yumer, E.; Wang, O.; Shechtman, E.; Lucey, S. ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018. doi:10.1109/cvpr.2018.00985.
47. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 11–18 December 2015. doi:10.1109/iccv.2015.425.
48. Song, S.; Yu, F.; Zeng, A.; Chang, A.X.; Savva, M.; Funkhouser, T. Semantic Scene Completion from a Single Depth Image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.28.
49. Litany, O.; Bronstein, A.; Bronstein, M.; Makadia, A. Deformable Shape Completion with Graph Convolutional Autoencoders. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018. doi:10.1109/cvpr.2018.00202.
50. Bogo, F.; Romero, J.; Pons-Moll, G.; Black, M.J. Dynamic FAUST: Registering Human Bodies in Motion. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.591.
51. Liu, X.; Qi, C.R.; Guibas, L.J. FlowNet3D: Learning Scene Flow in 3D Point Clouds. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. doi:10.1109/cvpr.2019.00062.
52. Mayer, N.; Ilg, E.; Hausser, P.; Fischer, P.; Cremers, D.; Dosovitskiy, A.; Brox, T. A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. doi:10.1109/cvpr.2016.438.
53. Mahapatra, D.; Antony, B.; Sedai, S.; Garnavi, R. Deformable medical image registration using generative adversarial networks. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018. doi:10.1109/isbi.2018.8363845.
54. Alipour, S.H.M.; Rabbani, H.; Akhlaghi, M.R. Diabetic Retinopathy Grading by Digital Curvelet Transform. *Comput. Math. Methods Med.* **2012**, *2012*, 1–11. doi:10.1155/2012/761901.
55. Radau, P.; Lu, Y.; Connelly, K.; Paul, G.; Dick, A.; Wright, G. Evaluation framework for algorithms segmenting short axis cardiac MRI. *MIDAS J. Cardiac MR Left Ventricle Segm. Chall.* **2009**, *49*.
56. Ofir, N.; Silberstein, S.; Levi, H.; Rozenbaum, D.; Keller, Y.; Bar, S.D. Deep Multi-Spectral Registration Using Invariant Descriptor Learning. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018 doi:10.1109/icip.2018.8451640.
57. Krizhevsky, A.; Hinton, G.; others. Learning multiple layers of features from tiny images. Technical report, 2009.

58. Brown, M.; Susstrunk, S. Multi-spectral SIFT for scene category recognition. In Proceedings of the CVPR, Providence, RI, USA, 20–25 June 2011. doi:10.1109/cvpr.2011.5995637.
59. Yan, P.; Xu, S.; Rastinehad, A.R.; Wood, B.J. Adversarial Image Registration with Application for MR and TRUS Image Fusion. In *MLMI*; Springer International Publishing: Cham, Switzerland, 2018; pp. 197–204. doi:10.1007/978-3-030-00919-9_23.
60. Wang, Z.; Rosa, S.; Yang, B.; Wang, S.; Trigoni, N.; Markham, A. 3D-PhysNet: Learning the Intuitive Physics of Non-Rigid Object Deformations. *arXiv* **2018**, arXiv:1805.00328.
61. Aoki, Y.; Goforth, H.; Srivatsan, R.A.; Lucey, S. PointNetLK: Robust & Efficient Point Cloud Registration Using PointNet. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. doi:10.1109/cvpr.2019.00733.
62. Chang, W.C.; Pham, V.T. 3-D Point Cloud Registration Using Convolutional Neural Networks. *Appl. Sci.* **2019**, *9*, 3273. doi:10.3390/app9163273.
63. Guan, S.; Meng, C.; Xie, Y.; Wang, Q.; Sun, K.; Wang, T. Deformable Cardiovascular Image Registration via Multi-Channel Convolutional Neural Network. *IEEE Access* **2019**, *7*, 17524–17534. doi:10.1109/access.2019.2894943.
64. Jack, D.; Pontes, J.K.; Sridharan, S.; Fookes, C.; Shirazi, S.; Maire, F.; Eriksson, A. Learning Free-Form Deformations for 3D Object Reconstruction. In *ACCV*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 317–333. doi:10.1007/978-3-030-20890-5_21.
65. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009. doi:10.1109/cvpr.2009.5206848.
66. Schaffert, R.; Wang, J.; Fischer, P.; Borsdorf, A.; Maier, A. Metric-Driven Learning of Correspondence Weighting for 2-D/3-D Image Registration. In *Lect. Notes Comput. Sci.*; Springer International Publishing: Cham, Switzerland, 2019; pp. 140–152. doi:10.1007/978-3-030-12939-2_11.
67. Smirnov, D.; Bessmeltsev, M.; Solomon, J. Deep Sketch-Based Modeling of Man-Made Shapes. *arXiv* **2019**, arXiv:1906.12337.
68. Yang, G.; Huang, X.; Hao, Z.; Liu, M.Y.; Belongie, S.; Hariharan, B. Pointflow: 3d point cloud generation with continuous normalizing flows. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 4541–4550.
69. Wang, Y.; Solomon, J.M. Deep Closest Point: Learning Representations for Point Cloud Registration. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019.
70. Wang, W.; Ceylan, D.; Mech, R.; Neumann, U. 3DN: 3D Deformation Network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019. doi:10.1109/cvpr.2019.00113.
71. Wang, L.; Fang, Y. Coherent point drift networks: Unsupervised learning of non-rigid point set registration. *arXiv* **2019**, arXiv:1906.03039.
72. Pais, G.D.; Ramalingam, S.; Govindu, V.M.; Nascimento, J.C.; Chellappa, R.; Miraldo, P. 3DRegNet: A Deep Neural Network for 3D Point Registration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 7193–7203.
73. Choi, S.; Zhou, Q.Y.; Koltun, V. Robust reconstruction of indoor scenes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015. doi:10.1109/cvpr.2015.7299195.
74. Zhou, B.; Lapedriza, A.; Xiao, J.; Torralba, A.; Oliva, A. Learning deep features for scene recognition using places database. In Proceedings of the Advances in Neural Information Processing Systems 27 (NIPS 2014), Montreal, MTL, Canada, 8–13 December 2014; pp. 487–495.
75. Li, X.; Pontes, J.K.; Lucey, S. Deterministic PointNetLK for Generalized Registration. *arXiv* **2020**, arXiv:2008.09527.
76. Yuan, W.; Eckart, B.; Kim, K.; Jampani, V.; Fox, D.; Kautz, J. DeepGMR: Learning Latent Gaussian Mixture Models for Registration. *arXiv* **2020**, arXiv:2008.09088.
77. Handa, A.; Whelan, T.; McDonald, J.; Davison, A.J. A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 1524–1531.

78. Zhang, X.; Jian, W.; Chen, Y.; Yang, S. Deform-GAN: An Unsupervised Learning Model for Deformable Registration. *arXiv* **2020**, arXiv:2002.11430.
79. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; others. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024.
80. Saval-Calvo, M.; Orts-Escolano, S.; Azorin-Lopez, J.; Garcia-Rodriguez, J.; Fuster-Guillo, A.; Morell-Gimenez, V.; Cazorla, M. Non-rigid point set registration using color and data downsampling. In Proceedings of the 2015 International Joint Conference on Neural Networks (IJCNN), Killarney, Ireland, 12–17 July 2015. doi:10.1109/ijcnn.2015.7280765.
81. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. doi:10.1038/nature14539.
82. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Van Esesn, B.C.; Awwal, A.A.S.; Asari, V.K. The history began from alexnet: A comprehensive survey on deep learning approaches. *arXiv* **2018**, arXiv:1803.01164.
83. Bench-Capon, T.J.M. Knowledge representation: an approach to artificial intelligence. *A.P.I.C. Series* **1990**, *32*, 220.
84. Norman, D.A. Some Observations on Mental Models. In *Human-Computer Interaction: A Multidisciplinary Approach*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1987; pp. 241–244.
85. Nersessian, N. How do scientists think?Capturing the dynamics of conceptual change in science. *Cogn. Model. Sci.* **1992**, *15*, 3–44.
86. Greca, I.M.; Moreira, M.A. Mental models, conceptual models, and modelling. *Int. J. Sci. Educ.* **2000**, *22*, 1–11. doi:10.1080/095006900289976.
87. Mahadevan, S. Imagination machines: A new challenge for artificial intelligence. In Proceedings of the AAAI, New Orleans, LA, USA, 2–7 February 2018.
88. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++:Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5099–5108.
89. Charles, R.Q.; Su, H.; Kaichun, M.; Guibas, L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. doi:10.1109/cvpr.2017.16.
90. Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for real-time object recognition. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015. doi:10.1109/iros.2015.7353481.
91. Xu, Y.; Fan, T.; Xu, M.; Zeng, L.; Qiao, Y. SpiderCNN: Deep Learning on Point Sets with Parameterized Convolutional Filters. In *ECCV*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 90–105. doi:10.1007/978-3-030-01237-3_6.
92. Lee, S.H.; Kim, H.U.; Kim, C.S. ELF-Nets: Deep Learning on Point Clouds Using Extended Laplacian Filter. *IEEE Access* **2019**, *7*, 156569–156581. doi:10.1109/access.2019.2949785.
93. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. doi:10.1109/cvpr.2016.90.
94. Jaderberg, M.; Simonyan, K.; Zisserman, A. Spatial transformer networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, MTL, Canada, 7–12 December 2015; pp. 2017–2025.
95. Shan, S.; Yan, W.; Guo, X.; Chang, E.I.; Fan, Y.; Xu, Y. Unsupervised end-to-end learning for deformable medical image registration. *arXiv* **2017**, arXiv:1711.08608.
96. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Dalca, A.V.; Guttag, J. An Unsupervised Learning Model for Deformable Medical Image Registration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018. doi:10.1109/cvpr.2018.00964.
97. Ronneberger, O.; Fischer, P.; Brox, T. U-Net:Convolutional Networks for Biomedical Image Segmentation. In *Lect. Notes Comput. Sci.*; Springer International Publishing: Cham, Switzerland, 2015; pp. 234–241. doi:10.1007/978-3-319-24574-4_28.

98. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient conv. neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
99. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In *IJCAI*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1981; IJCAI'81, pp. 674–679.
100. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic Graph CNN for Learning on Point Clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. doi:10.1145/3326362.
101. Lu, H.; Shi, H. Deep Learning for 3D Point Cloud Understanding: A Survey. *arXiv preprint arXiv:2009.08920* **2020**.