applied sciences

MDPI

Article

# Formula-Driven Supervised Learning in Computer Vision: A Literature Survey

**Abdul Mueed Hafiz [1], Mahmoud Hassaballah [2,3,*] and Adel Binbusayyis [4]**

1 Department of Electronics & Communication Engineering, Institute of Technology, University of Kashmir, Srinagar 190006, J&K, India
2 Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, AlKharj 16278, Saudi Arabia
3 Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena 83523, Egypt
4 Department of Software Engineering, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, AlKharj 16278, Saudi Arabia
* Correspondence: mah.ali@psau.edu.sa

**Abstract:** Current computer vision research uses huge datasets with millions of images to pre-train vision models. This results in escalation of time and capital, ethical issues, moral issues, privacy issues, copyright issues, fairness issues, and others. To address these issues, several alternative learning schemes have been developed. One such scheme is formula-based supervised learning (FDSL). It is a form of supervised learning, which involves the use of mathematically generated images for the pre-training of deep models. Promising results have been obtained for computer-vision-related applications. In this comprehensive survey paper, a gentle introduction to FDSL is presented. The supporting theory, databases, experimentation and ensuing results are discussed. The research outcomes, issues and scope are also discussed. Finally, some of the most promising future directions for FDSL research are discussed. As FDSL is an important learning technique, this survey represents a useful resource for interested researchers working on solving various problem in computer vision and related areas of application.

**Keywords:** formula-driven supervised learning; fractals; deep learning; visual transformers; ViTs; CNNs; object recognition; computer vision

check for updates

**Citation:** Hafiz, A.M.; Hassaballah, M.; Binbusayyis, A. Formula-Driven Supervised Learning in Computer Vision: A Literature Survey. *Appl. Sci.* **2023**, *13*, 723. https://doi.org/10.3390/app13020723

Academic Editor: Andrea Prati

Received: 17 November 2022
Revised: 13 December 2022
Accepted: 15 December 2022
Published: 4 January 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Deep learning is a powerful approach for performing different types of computer vision tasks [1], such as object recognition [2–5], image segmentation [6–9], visual captioning [10], etc. Deep learning also enables other tasks, such as natural language processing (NLP) [11]. Deep networks pre-trained on large image datasets, e.g., ImageNet [12] have been used after fine-tuning for two important reasons [13]: First, the features learned by deep networks from large datasets help deep networks to generalize more effectively and rapidly; Second, pre-trained deep networks are successful at avoiding over-fitting during fine-tuning for smaller downstream tasks.

It is well known that the performance of deep networks depends on their architecture as well as their training [14–17]. A multitude of successful deep networks have been developed with a large number of parameters. To train these parameters, a very large number of training images are required; hence, the need for large-scale image datasets. Popular deep networks include AlexNet [18], VGG [19], GoogLeNet [20], ResNet [21], and DenseNet [22]. Popular large-scale image datasets include ImageNet [12] and OpenImage [23]. Deep networks have achieved state-of-the-art performance for many computer-vision applications [2,6,10,24–28].

In spite of their success, the training of deep networks has become expensive and time-consuming. This is due to the laborious collection and manual annotation of the high volume of data required with large-scale datasets. For example, ImageNet [12], which is a popular dataset, has about 1.3 M annotated images with 1 K classes. In ImageNet, every image has been manually annotated and the annotation process is, therefore, substantial. In addition to this, collecting and annotating video data is more expensive still due to its temporal aspect. For example, the Kinetics video dataset [29] includes 500 K human action videos with 600 classes. Every video in the dataset is around 10 seconds long. Several Turkish Amazon workers were involved in the collection and annotation of this large-scale dataset.

Bearing the above issues in mind, it is necessary to avoid lengthy and expensive manual annotation. To this end, different learning schemes have been proposed for training deep networks. These schemes include semi-supervised learning (SSL) [30–39], weakly supervised learning (WSL) [40–44], unsupervised learning (USL) [45–57], and self-supervised learning [58–76]. The advantage of these schemes is that they can be used to train deep networks without the use of labeled data; hence, the expensive manual annotation process can be avoided. This can save time as well as expense. A promising supervised learning scheme, which involves use of synthetic images with auto-generated labels, is formula-driven supervised learning (FDSL) [77]. This learning scheme automates the dataset creation process and has produced promising results when used for pre-training. FDSL represents a viable alternative to other learning schemes which offer training using unlabeled data. FDSL has the potential to alter the way in which large-data models are trained, i.e., without manual collection or manual annotation. This potentially critical development, opening a new area of research, motivated us to conduct a literature survey on FDSL. It is hoped that, through such surveys, improved strategies for addressing the critical issues associated with the training of large-data models will emerge and become widely applied.

The main contributions of the paper can be summarized as follows:

- A detailed review of recent supervised learning schemes with respect to formula-driven supervised learning is presented.
- Extensive manual data collection and annotation methods for training large-data models based on FDSL are discussed using synthetic datasets.
- The state-of-the-art, the advantages and disadvantages, and the limitations and future potential of the techniques are considered.

The remainder of the paper is structured as follows: In Section 2, we discuss different deep-learning schemes. This is followed by Section 3, wherein, we discuss formula-driven supervised learning. Section 4 considers issues associated with FDSL and its future potential. We present our conclusions in Section 5.

## 2. Deep-Learning Schemes

In this section, we discuss different deep-learning schemes which are grouped into four categories: supervised, semi-supervised, weakly supervised, and unsupervised learning.

### 2.1. *Brief Discussion*

#### 2.1.1. Supervised Learning

In supervised learning, for a dataset $X$ containing data denoted by $X_i \in X$, for every data entry there is a manually annotated label $Y_i$. If there are $N$ labels for the training set $D = \{X_i\}_{i=0}^{N}$, the loss function is given by:

$$loss(D) = \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} loss(X_i, Y_i) \tag{1}$$

where $\theta$ is defined as the model convergence error parameter in the loss equation.

When trained accurately with manually annotated labels, supervised learning schemes achieve state-of-the-art performance on various computer vision tasks [2,6,18,26]. In spite of this, there are burdens of collection of data and investment of the time and effort required for manual annotation, which, in turn, requires specific skills. To address these issues, semi-supervised, weakly supervised and unsupervised learning schemes have been proposed.

#### 2.1.2. Semi-Supervised Learning

In the semi-supervised learning scheme [30–39], for a small labeled dataset $X$ and a large unlabeled dataset $Z$, for every data $X_i \in X$, there exists a manually annotated label $Y_i$. Given $N$ labels in the training set $D_1 = \{X_i\}_{i=0}^{N}$ and $M$ unlabeled training set $D_2 = \{Z_i\}_{i=0}^{N}$, the loss function is given by:

$$loss(D_1, D_2) = \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} loss(X_i, Y_i) + \frac{1}{M} \sum_{i=1}^{M} loss(Z_i, R(Z_i, X)) \tag{2}$$

where $R(Z_i, X)$ is a task-specific function relating every unlabeled data $Z_i$ with the labeled training set $X$. Semi-supervised learning can be helpful for a very large corpus of data where manual annotation is not extensive. However, its accuracy is not as good as that of traditional approaches.

#### 2.1.3. Weakly Supervised Learning

In the weakly supervised learning scheme [40–44], for a dataset $X$ having data $X_i \in X$, there exists a coarse-grained label $C_i$. For a training set $D_i = \{X_i\}_{i=0}^{N}$, the loss function is given by:

$$loss(D) = \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} loss(X_i, C_i) \tag{3}$$

The costs associated with weakly supervised learning are much less than for supervised learning; hence, large sparsely labeled datasets are much easier to build. Many studies have proposed learning from images collected from the Internet with the use of hashtag labels [78,79]. Good performance has been observed after using these techniques, although, again, the performance may not be as good as that using traditional methods. Improving the efficacy of the approach requires further research.

#### 2.1.4. Unsupervised Learning

In unsupervised learning [80–83], manually annotated labels are not used. In fact, these techniques do not use labels. Formulas can serve as effective model representations [84]. The deep networks are trained using auto-generated pseudo-labels; hence, there is no need for manual annotation. One type of unsupervised learning is the self-supervised learning scheme. Several self-supervised learning schemes have been proposed [58–76]. This type of learning scheme is referred to in some reports as unsupervised learning [45–57]. In comparison to supervised learning schemes, which require data paired with labels in the form $(X_i, Y_i)$, where $Y_i$ is the manually annotated label, self-supervised learning uses pseudo-label data pairs in the form $(X_i, P_i)$. Here the pseudo-label $P_i$ is auto-generated for the task without using any manual annotation. The pseudo-labeling is performed with the help of image attributes, such as the image context [45,85–87].

For a training set $D_i = \{P_i\}_{i=0}^{N}$ having $N$ labels, the loss function is given by:

$$loss(D) = \min_{\theta} \frac{1}{N} \sum_{i=1}^{N} loss(X_i, P_i) \tag{4}$$

Recently, unsupervised learning schemes have become quite popular. However, the complexity of the process is greater than that of competing techniques, and performance might not necessarily be equivalent to the latter.

### 2.2. Recent Trends in Deep-Learning Schemes

Here, we indicate some recent trends in the development and application of deep-learning schemes referred to above. Figure 1 shows the number of related global publications over the past 10 years. Figure 2 shows the total number of related global publications (with breakdown) over the last 10 years. Figure 3 presents a breakdown of the numbers of related global publications for 2010 and 2021, respectively. In the figures, SL, SSL, WSL, and USL stand for supervised learning, semi-supervised learning, weakly supervised learning, and unsupervised learning, respectively.

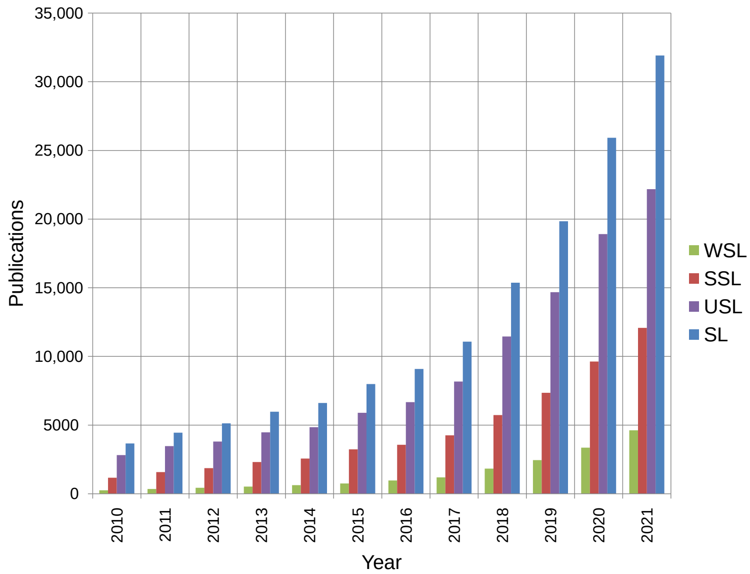


**Figure 1.** Related global publications for past 10 years. The trends indicate a notable increase in the number of global publications. Supervised learning (SL) and unsupervised learning (USL) publications tend to be equally dominant.

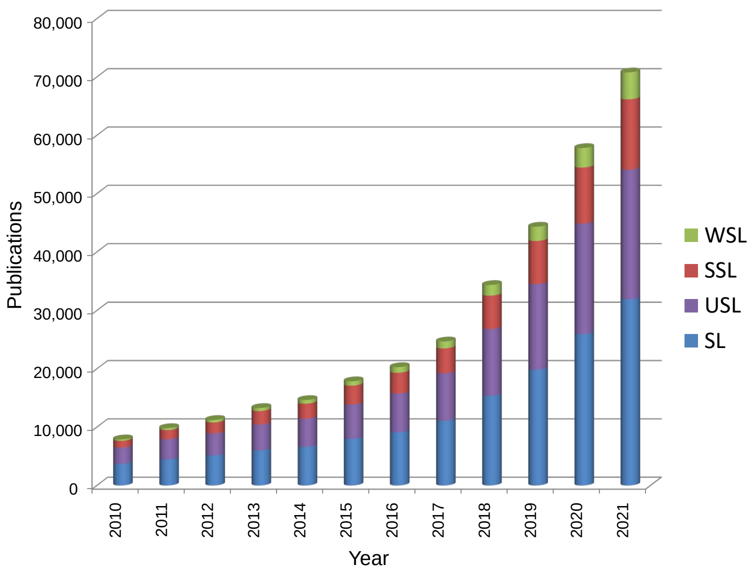


**Figure 2.** Total related global publications with breakdown for past 10 years. The trends indicate a notably increased dominance of supervised learning (SL) and unsupervised learning (USL) to the same degree.

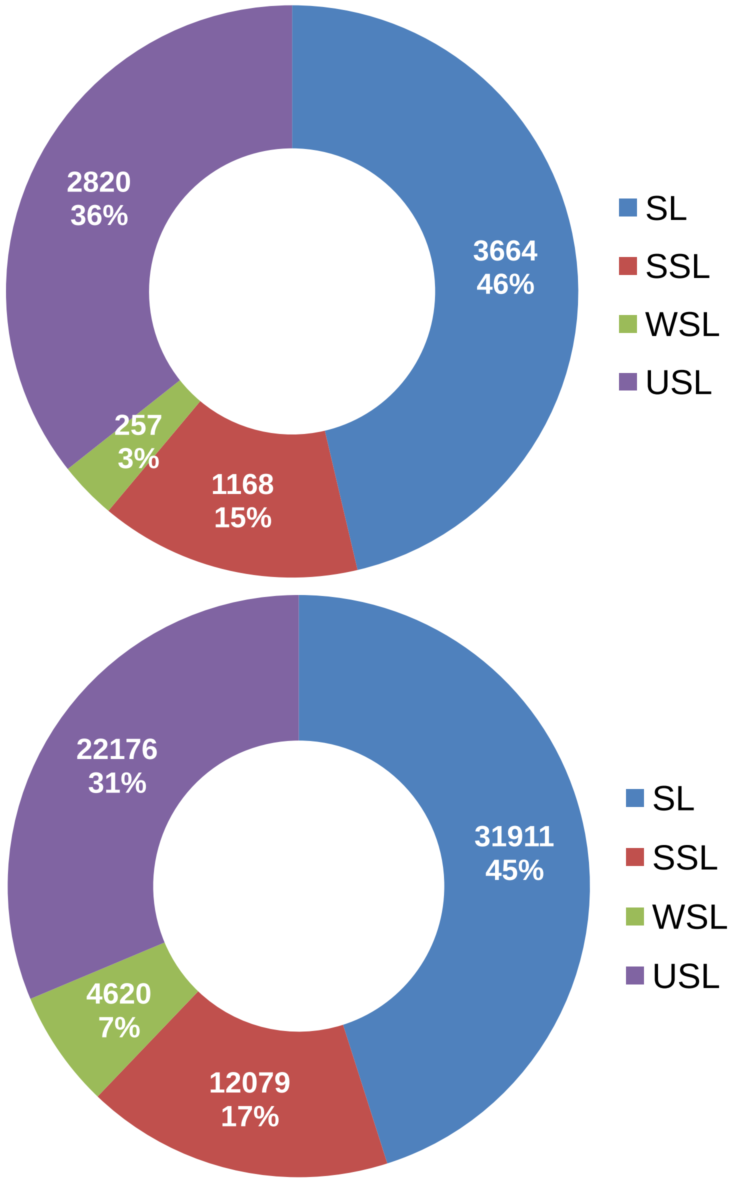


**Figure 3.** Breakdown of related global publications for the years (**Top**) 2010 and (**Bottom**) 2021. For the two sample years, there was a notable increase in the number of global publications relating to supervised learning (SL) and unsupervised learning (USL).

In the next section, we discuss the formula-driven supervised learning scheme, which, as the name indicates, is a supervised learning scheme.

## 3. Formula-Driven Supervised Learning

The development of FDSL [77] was motivated by the need to find a technique for the automatic generation of pre-training datasets without taking images from nature. The authors who proposed the concept believed that FDSL would outperform other pre-training techniques by being more fair, private, and ethical. They also believed that it would reduce the burden of manual annotation and the massive downloading of images.

### *3.1. Background of FDSL*

Current computer-vision deep-learning schemes use image datasets that have millions of images to train visual architectures. Although outstanding results have been achieved using such schemes, there are serious issues associated with them. These include a huge burden of manual annotation, the cost of image collection and labeling, privacy issues, copyright issues, ethical issues, and fairness concerns [77]. In [77], the authors propose pre-training of computer-vision models without the use of natural images to overcome these issues. They refer to their technique as formula-driven supervised learning (FDSL). The technique is used for the creation of pairs of images and labels using mathematical formulae.

FDSL can be mathematically expressed by

$$\arg\max_{M} (\mathbb{E}_{y,s}[l(M(x), y)] \textbf{ s.t. } x = F(\theta, s), y = \theta) \tag{5}$$

In the above equation, $\mathbb{E}$ is the Euclidean space representation of the fractal, $M$ is the classification network used for pre-training, $l$ is the classification-loss, $x$ is the image obtained by generation, and $y$ is the image label. The FDSL images are mathematically synthesized using a formula $F$ with a parameter $\theta$, which is an affine transformation parameter set related to shift or rotation, and a randomly generated seed $s$. The aim of the FDSL training is prediction of $\theta$, using which the image $x$ was generated. It is assumed that the label $y$ has a uniformly distributed value on a set of discrete values $\Theta = \{\theta_k\}_k^K$. This feature introduces a classification-loss $l$ over $K$ classes, e.g., a cross-entropy-based loss function. Figure 4 presents an overview of FDSL.

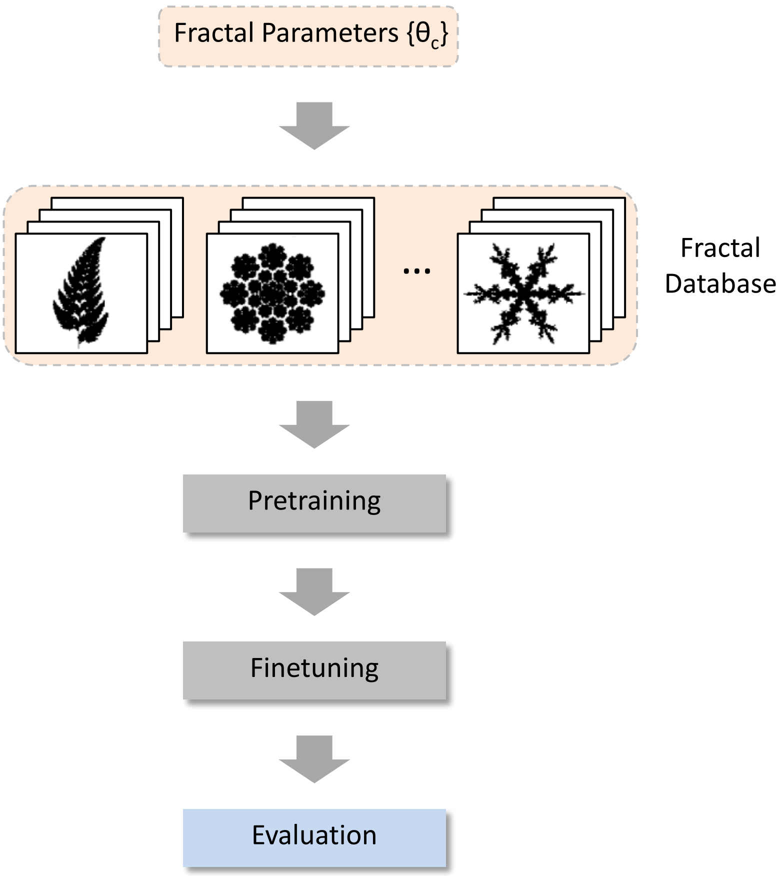


**Figure 4.** Overview of the FDSL technique using a fractal-based database [77].

### 3.2. Learning Frameworks

Currently, supervised learning represents the state-of-the-art in computer vision [18–21,88–91]. Research has recently been undertaken to decrease the data volume for unsupervised, weakly supervised and self-supervised training to avoid the need for manual annotation. Self-supervised training has the potential to create pre-trained architectures cost-efficiently. This involves use of a basic, but relevant, 'pre-text task' [45,50,51,69,87,92]. Although earlier techniques [45,51,87] were not suitable as alternatives to manual annotation, new techniques, such as SimCLR [93], MoCo [94], and DeepCluster [53], are much better. Semi-supervised learning (SSL) [30–39] has the potential to replace human annotation, although there are significant issues with downloading, privacy and fairness. FDSL [77] is superior to these techniques because it generates new mathematically formulated images along with their respective labels.

### 3.3. Formula Based Projection of Images

Fractals are one of the most popular mathematical image projection techniques. Fractal theory research is extensive [95–97]. It is used to render an image pattern using a basic mathematical expression [98–100] and to reconstruct architectures for object recognition [101–104]. Although rendering a fractal pattern leads to potential loss of its infinite representation for 2D images, humans naturally recognize such renderings. Since fractals occur naturally [95,105], the founders of FDSL claim [77] that fractals may aid in the learning of natural-image-based scenes and objects. They also consider [77] other techniques, such as Bezier curves [106] and Perlin noise [107] for rendering purposes. These techniques have been implemented and evaluated experimentally [77].

### 3.4. FDSL Datasets

As work on FDSL has increased, some interesting databases have been developed. Their details are provided below.

#### 3.4.1. Fractal DataBase

The fractal dataBase (FractalDB) [77] was developed for FDSL. It contains pairs of fractal images $I$ and their respective category labels $c$ [98], which are generated using an iterated function system (IFS) [98]. The IFS is defined over a metric space $\chi$ as:

$$\mathrm{IFS} = \{\chi; w_1, w_2, \cdots, w_N; p_1, p_2, \cdots, p_N\}, \tag{6}$$

where $w_i : \chi \to \chi$ is the transformation function, $p_i$ is probability with summation 1, and $N$ is the aggregate of transformations.

In IFS, each fractal $S = \{x_t\}_{t=0}^{\infty} \in \chi$ is randomly constructed [98] using a two-step algorithm by repeating it for $t = 0, 1, 2, \cdots$ from a starting coordinate $\mathbf{x}_0$. First, predefined probabilities $p_i = p(w^* = w_i)$ are used to select a transformation $w^*$ from $\{w_1, \cdots, w_N\}$ for determining the $i^{th}$ transformation. Next, a fresh point $\boldsymbol{x}_{t+1} = w^*(\boldsymbol{x}_t)$ is generated.

2D fractals are constructed using a Euclidean space $\chi = \mathbb{R}^2$ by an affine transformation [98]. The transformation has six parameters $\theta_i = (a_i, b_i, c_i, d_i, e_i, j_i)$ relating to rotation or shift:

$$w_i(\boldsymbol{x}; \theta_i) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} e_i \\ f_i \end{bmatrix} \tag{7}$$

The fractal image is generated by drawing dots on a uniform background.

IFS has a set of parameters along with their probabilities, given by:

$$\Theta = \{(\theta_i, p_i)\}_{i=1}^{N} \tag{8}$$

The authors of FDSL assume that every category has a unique $\Theta$. They generate 1k and 10k random categories for the FractalDB-1k and the FractalDB-10k datasets, respectively. In these datasets, $N$ is chosen from the distribution $\mathbb{N} = \{2, 3, 4, 5, 6, 7, 8\}$. $\theta_i$ has bounds $[-1, 1]$ for $i = 1, 2, \cdots, N$. $p_i$ is of the form:

$$p_i = \frac{\det(A_i)}{\sum_{i=1}^{N} \det(A_i)} \tag{9}$$

where $A_i = (a_i, b_i, c_i, d_i)$ is the affine rotation. Finally, the new category $\Theta = \{(\theta_i, p_i)\}_{i=1}^{N}$ is accepted after further inspections.

Figure 5 shows fractal 2D images from the FractalDB dataset.

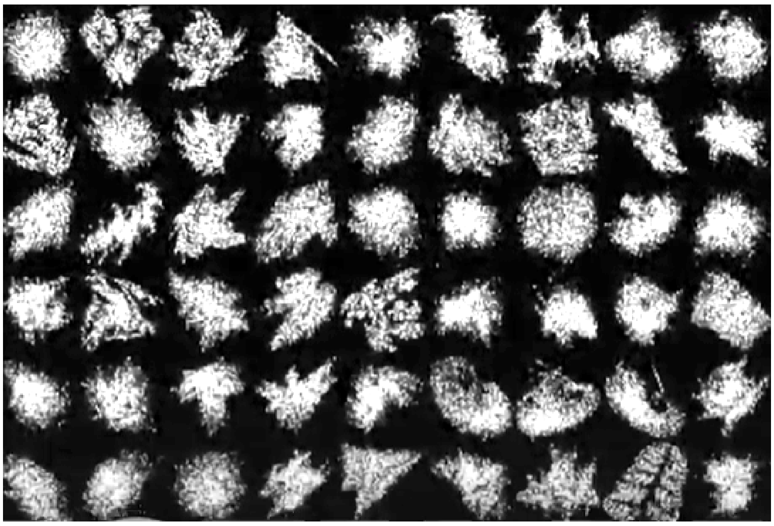

**Figure 5.** Some images from FractalDB [77].

#### 3.4.2. MV-FractalDB

Moving beyond FractalDB, which is a 2D image database, the authors of [108] developed an autogenerated multi-view image dataset for FDSL. They used fractal geometry to construct the dataset. The dataset has been named the multi-view fractal database (MV-FractalDB). Figure 6 shows some fractal images (3D) from the database.

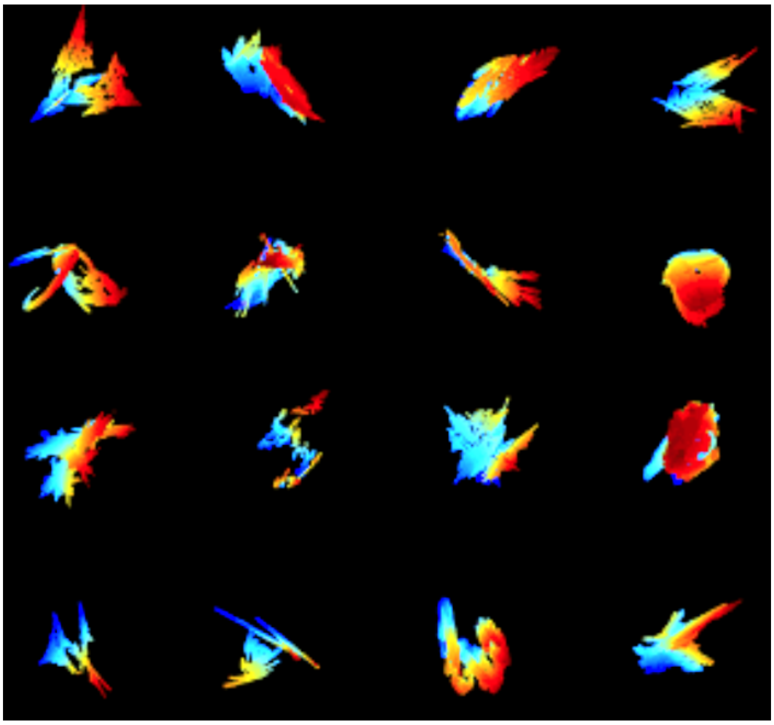

**Figure 6.** Sample images from MV-FractalDB [108].

MV-FractalDB has been used for pre-training deep models and promising results have been obtained. Based on experimentation [108], the MV-FractalDB pre-trained deep models were found to perform better than other self-supervised methods, such as SimCLR and MoCo. In addition, the MV-FractalDB pre-trained deep models converged faster than those trained on ImageNet. A performance comparison of MV-FractalDB pre-trained models against other state-of-the-art models is shown in Table 1.

**Table 1.** Performance in terms of %age accuracy on the ModelNet [109] and the MIRO [110,111] datasets, respectively [108]. (Note: SFSL = self-supervised learning, SL = supervised learning).

| Pretraining Technique | Learning Method | ModelNet40 (12 Views) | ModelNet40 (20 Views) | MIRO (20 Views) |
|---|---|---|---|---|
| — | — | 84.0 | 91.5 | 91.7 |
| ImageNet | SL | 88.1 | 96.1 | 100.0 |
| SimCLR | SFSL | 88.1 | 95.1 | 100.0 |
| MoCo | SFSL | 86.4 | 95.3 | 100.0 |
| FractalDB1k | FDSL | 87.4 | 94.9 | 100.0 |
| MV-FractalDB1k | FDSL | 87.6 | 95.7 | 100.0 |

3.4.3. Other Notable FDSL Databases

Another recent FDSL dataset is TileDB [112], which contains patterns made with tiles. A tile is a group of wallpapers with 2D repetition and complicated textures. It is obtained by adding three operations for hexagonally shaped tiles: (i) moving vertices, (ii) deforming edges, and (iii) moving symmetrically in a specific direction. Using this basic technique, the TileDB dataset was created with 1000 classes and 1071 images for each class. The FractalDB, Kataoka et al. [91] proposed some formula-based datasets for the pretraining of computer-vision models, such as convolutional neural networks [14–16,113,114] and visual transformers [17,115]. These are the Perlin noise-based *PerlinNoiseDB* and the Bezier curve-based *BezierCurveDB*. The DeiT model [116] was pretrained and fine-tuned using these formula-based image datasets. This enabled determination that FractalDB was the best choice for FDSL of computer-vision models developed to date. This is confirmed by data presented in Table 2. Improvements in the percentage classification accuracy of up to +18.5, +23.9, +74.4, +21.2 on the Cifar-10, Cifar-100, Cars dataset, and Flowers dataset, respectively, using FractalDB-1k were observed.

**Table 2.** Performance comparison of visual transformer pretraining in terms of % age accuracy with FractalDB1k and other FDSL datasets for BezierCurveDB and PerlinNoiseDB [77].

| Dataset | Cifar-10 | Cifar-100 | Cars | Flowers |
|---|---|---|---|---|
| — | 78.3 | 57.7 | 11.6 | 77.1 |
| PerlinNoiseDB | 94.5 | 77.8 | 62.3 | 96.1 |
| BeizerCurveDB | 96.7 | 80.3 | 82.8 | 98.5 |
| FractalDB1k | 96.8 | 81.6 | 86.0 | 98.3 |

The authors of FractalDB [77] investigated various adaptable parameters, including #category, #instance, filling rate, fractal weight, #dot, and image size. For further information about these parameters, readers may refer to [77]. The experimentation results reported in [77] are shown in Table 3.

As it can be seen from Table 3, FDSL showed notable performance improvements when using FractalDB for pre-training and its potential was highlighted. From the experimental results presented in Tables 1–3, the performance of FDSL was generally equivalent to that of other competing methods. This performance was achieved using different models pretrained on synthetic data, in contrast to the huge datasets created by conventional methods.

**Table 3.** Performance comparison in terms of % age accuracy of FractalDB1k [77], FractalDB10k [77], TileDB [112], DeepCluster10k (DC) [53], ImageNet100 [12], ImageNet1k [12], Places30 [117] and Places365 [117] on pretrained ResNet-50 for various datasets, as given in [77]. The datasets used were CIFAR10 (C-10) [118], CIFAR100 (C-100) [118], ImageNet1k (ImNt-1k) [12], Places365 (P-365) [117], PascalVOC-2012 (VOC-12) [119] and Omniglot (OG) [120]. (Note: SFSL=self-supervised learning, SL=supervised learning).

| Technique | Image Type | Method | C-10 | C-100 | ImNt-1k | P-365 | VOC-12 | OG |
|---|---|---|---|---|---|---|---|---|
| — | – | – | 87.6 | 62.7 | 76.1 | 49.9 | 58.9 | 1.1 |
| DC | Natural | SFSL | 89.9 | 66.9 | 66.2 | 51.5 | 67.5 | 15.2 |
| Places30 | Natural | SL | 90.1 | 67.8 | 69.1 | – | 69.5 | 6.4 |
| Places365 | Natural | SL | 94.2 | 76.9 | 71.4 | – | 78.6 | 10.5 |
| ImageNet100 | Natural | SL | 91.3 | 70.6 | – | 49.7 | 72.0 | 12.3 |
| ImageNet1k | Natural | SL | 96.8 | 84.6 | – | 50.3 | 85.8 | 17.5 |
| TileDB | Synthetic | FDSL | 92.5 | 73.7 | – | – | 71.4 | – |
| FractalDB1k | Synthetic | FDSL | 93.4 | 75.7 | 70.3 | 49.5 | 58.9 | 20.9 |
| FractalDB10k | Synthetic | FDSL | 94.1 | 77.3 | 71.5 | 50.8 | 73.6 | 29.2 |

## 4. Issues and Future Scope

### *4.1. Issues*

Although FDSL is promising, there are issues associated with it. These include the limited number of FDSL formulae and their limited parameters [77]. This issue impacts on pre-training and final validation, leading to lower performance compared to natural-image-based pre-training, as is shown in Tables 1–3, where the performance is seen to be lower. Capturing the richness of natural images using mathematically generated images remains an issue. Capturing color variations and naturally occurring patterns, textures, etc., as found in natural image pretraining, is also an issue. The limitations in the number of mathematically generated patterns may also affect performance [77]. FDSL currently lags behind other competing techniques due to its semi-simplistic model approximation of the mathematical formulae [77]. Richer mathematical models can help build better pseudo-natural data, which can be even richer and more compact than natural data. This is, again, subject to experimental confirmation.

### *4.2. Future Scope*

In spite of the issues associated with FDSL, it is hoped that, with more in-depth research, better results will be achieved. Colored fractals are also available [108], leading to better generalizations. Developing stronger mathematical formula for automatic image generation is a potential area of interest. Pattern generation identical to that for natural objects is another potential area of interest. In addition, moving beyond simple fractal and basic patterns [77] by generating richer artificial images can lead to much better results. The combination of FDSL with other training techniques is also a promising research area, which may lead to enhanced performance. With the development of stronger large-data models, e.g., visual transformers [17,115], FDSL performance can improve. Moreover, by developing large-data models, which are more specifically adaptable to mathematical models, more success may be achieved for FDSL as it relies on mathematical modeling.

## 5. Conclusions

In this survey paper, a gentle introduction to formula-driven supervised learning (FDSL) has been provided. Various aspects of FDSL were discussed including its mathematical background, methodology, experimental results, issues and future scope. It was observed that FDSL produced promising results for object recognition. It was also emphasized that FDSL addresses the issues associated with natural image pre-training on huge datasets making it a suitable candidate for computer-vision applications. These issues include manual annotation costs and time, privacy, and fairness. It is hoped that the readers will be encouraged to learn about and undertake research in this interesting and promising area of computer vision.

**Author Contributions:** Conceptualization, A.M.H. and M.H.; methodology, A.M.H. and M.H.; resources, A.B. and M.H.; writing—original draft preparation, A.M.H.; writing—review and editing, M.H. and A.B.; funding acquisition, M.H. and A.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

**Acknowledgments:** This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1444).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hassaballah, M.; Awad, A.I. *Deep Learning in Computer Vision: Principles and Applications*; CRC Press: Boca Raton, FL, USA, 2020.
2. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23-28 June 2014; pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.
3. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. https://doi.org/10.1109/ICCV.2015.169.
4. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
5. Sajid, F.; Javed, A.R.; Basharat, A.; Kryvinska, N.; Afzal, A.; Rizwan, M. An Efficient Deep Learning Framework for Distracted Driver Detection. *IEEE Access* **2021**, *9*, 169270–169280. https://doi.org/10.1109/ACCESS.2021.3138137.
6. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. https://doi.org/10.1109/TPAMI.2016.2572683.
7. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184.
8. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 21–26 July 2017; pp. 6230–6239. https://doi.org/10.1109/CVPR.2017.660.
9. Hafiz, A.M.; Bhat, G.M. A survey on instance segmentation: state of the art. *Int. J. Multimed. Inf. Retr.* **2020**, *9*, 171–189.
10. Vinyals, O.; Toshev, A.; Bengio, S.; Erhan, D. Show and tell: A neural image caption generator. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE Computer Society: Los Alamitos, CA, USA, 2015; pp. 3156–3164. https://doi.org/10.1109/CVPR.2015.7298935.
11. Amanat, A.; Rizwan, M.; Javed, A.R.; Abdelhaq, M.; Alsaqour, R.; Pandya, S.; Uddin, M. Deep Learning for Depression Detection from Textual Data. *Electronics* **2022**, *11*, 676. https://doi.org/10.3390/electronics11050676.
12. 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; pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
13. Jing, L.; Tian, Y. Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 4037–4058. https://doi.org/10.1109/TPAMI.2020.2992393.
14. Hafiz, A.M.; Bhat, G.M. Deep Network Ensemble Learning applied to Image Classification using CNN Trees. *arXiv* **2020**. https://doi.org/10.48550/arXiv.2008.00829.
15. Hafiz, A.M.; Hassaballah, M. Digit Image Recognition Using an Ensemble of One-Versus-All Deep Network Classifiers. In Proceedings of the Information and Communication Technology for Competitive Strategies (ICTCS 2020), Jaipur, Rajasthan, India, 11–12 December 2020; Kaiser, M.S., Xie, J., Rathore, V.S., Eds.; Springer: Singapore, 2021; pp. 445–455.
16. Hafiz, A.M.; Bhat, G.M., Fast training of deep networks with one-class CNNs. In *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough: Latest Trends in AI, Volume 2*, 27 April 2021; Gunjan, V.K., Zurada, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 409–421. https://doi.org/10.1007/978-3-030-68291-0_33.
17. Hafiz, A.M.; Parah, S.A.; Bhat, R.U.A. Attention mechanisms and deep learning for machine vision: A survey of the state of the art. *arXiv* **2021**. https://doi.org/10.48550/arXiv.2106.07550.
18. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 3–6 December 2012; Pereira, F., Burges, C., Bottou, L., Weinberger, K., Eds.; Curran Associates, Inc.: Red Hook, NY, USA 2012; Volume 25.

19. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**. https://doi.org/10.48550/arXiv.1409.1556.
20. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; IEEE Computer Society: Los Alamitos, CA, USA, 2015; pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA, 26 June–1 July 2016; pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
22. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 22–25 July 2017; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.
23. Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Kolesnikov, A.; et al. The open images dataset v4. *Int. J. Comput. Vis.* **2020**, *128*, 1956–1981.
24. Hassaballah, M.; Khalid M., H. *Recent Advances in Computer Vision: Theories and Applications*; Springer: Berlin/Heidelberg, Germany, 2019.
25. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 22–25 July 2017.
26. Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; Paluri, M. Learning Spatiotemporal Features With 3D Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
27. Hafiz, A.M.; Bhat, G.M. A Survey of Deep Learning Techniques for Medical Diagnosis. In Proceedings of the Information and Communication Technology for Sustainable Development, Goa, India, 23–24 Jul, 2020; Tuba, M., Akashe, S., Joshi, A., Eds.; Springer: Singapore, 2020; pp. 161–170.
28. Hafiz, A.M.; Bhat, R.U.A.; Parah, S.A.; Hassaballah, M. SE-MD: A Single-encoder multiple-decoder deep network for point cloud generation from 2D images. *arXiv* **2021**. https://doi.org/10.48550/arXiv.2106.15325.
29. Kay, W.; Carreira, J.; Simonyan, K.; Zhang, B.; Hillier, C.; Vijayanarasimhan, S.; Viola, F.; Green, T.; Back, T.; Natsev, P.; et al. The Kinetics Human Action Video Dataset. *arXiv* **2017**. https://doi.org/10.48550/arXiv.1705.06950.
30. Yang, X.; Song, Z.; King, I.; Xu, Z. A Survey on Deep Semi-supervised Learning. *arXiv* **2021**. https://doi.org/10.48550/arXiv.2103.00550.
31. Van Engelen, J.E.; Hoos, H.H. A survey on semi-supervised learning. *Mach. Learn.* **2020**, *109*, 373–440.
32. Li, Y.F.; Liang, D.M. Safe semi-supervised learning: a brief introduction. *Front. Comput. Sci.* **2019**, *13*, 669–676.
33. Yalniz, I.Z.; Jégou, H.; Chen, K.; Paluri, M.; Mahajan, D. Billion-scale semi-supervised learning for image classification. *arXiv* **2019**. https://doi.org/10.48550/arXiv.1905.00546.
34. Sohn, K.; Zhang, Z.; Li, C.L.; Zhang, H.; Lee, C.Y.; Pfister, T. A Simple Semi-Supervised Learning Framework for Object Detection. *arXiv* **2020**. https://doi.org/10.48550/arXiv.2005.04757.
35. Jeong, J.; Lee, S.; Kim, J.; Kwak, N. Consistency-based Semi-supervised Learning for Object detection. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA 2019; Volume 32.
36. Jiang, B.; Zhang, Z.; Lin, D.; Tang, J.; Luo, B. Semi-Supervised Learning With Graph Learning-Convolutional Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
37. Oliver, A.; Odena, A.; Raffel, C.A.; Cubuk, E.D.; Goodfellow, I. Realistic Evaluation of Deep Semi-Supervised Learning Algorithms. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, Canada, 8–14 December 2018; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA 2018; Volume 31.
38. Li, Q.; Han, Z.; Wu, X.M. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence—AAAI'18/IAAI'18/EAAI'18, New Orleans Riverside, New Orleans, USA, 2–7 February 2018.
39. Feng, W.; Zhang, J.; Dong, Y.; Han, Y.; Luan, H.; Xu, Q.; Yang, Q.; Kharlamov, E.; Tang, J. Graph Random Neural Networks for Semi-Supervised Learning on Graphs. In Proceedings of the Advances in Neural Information Processing Systems (Virtual-only Conference), 6–12 December 2020; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: Red Hook, NY, USA 2020, Volume 33, pp. 22092–22103.
40. Saito, S.; Yang, J.; Ma, Q.; Black, M.J. SCANimate Weakly Supervised Learning of Skinned Clothed Avatar Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 2886–2897.
41. Li, Y.F.; Guo, L.Z.; Zhou, Z.H. Towards Safe Weakly Supervised Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 334–346. https://doi.org/10.1109/TPAMI.2019.2922396.

42. Ahn, J.; Cho, S.; Kwak, S. Weakly Supervised Learning of Instance Segmentation With Inter-Pixel Relations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
43. Zhang, M.; Zhou, Y.; Zhao, J.; Man, Y.; Liu, B.; Yao, R. A survey of semi-and weakly supervised semantic segmentation of images. *Artif. Intell. Rev.* **2020**, *53*, 4259–4288.
44. Baldassarre, F.; Smith, K.; Sullivan, J.; Azizpour, H. Explanation-Based Weakly-Supervised Learning of Visual Relations with Graph Networks. In Proceedings of the Computer Vision—ECCV 2020 (Online Conference), 23–28 August 2020; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 612–630.
45. Gidaris, S.; Singh, P.; Komodakis, N. Unsupervised Representation Learning by Predicting Image Rotations. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1803.07728.
46. Srivastava, N.; Mansimov, E.; Salakhutdinov, R. Unsupervised Learning of Video Representations Using LSTMs. In Proceedings of the 32nd International Conference on International Conference on Machine Learning—ICML'15, Lille, France, 6–11 July 2015; Volume 37, pp. 843–852.
47. Wang, X.; Gupta, A. Unsupervised Learning of Visual Representations using Videos. *arXiv* **2015**. https://doi.org/10.48550/arXiv.1505.00687.
48. Lee, H.Y.; Huang, J.B.; Singh, M.; Yang, M.H. Unsupervised Representation Learning by Sorting Sequences. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
49. Misra, I.; Zitnick, C.L.; Hebert, M. Shuffle and Learn: Unsupervised Learning Using Temporal Order Verification. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 527–544.
50. Doersch, C.; Gupta, A.; Efros, A.A. Unsupervised Visual Representation Learning by Context Prediction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
51. Zhang, R.; Isola, P.; Efros, A.A. Split-Brain Autoencoders: Unsupervised Learning by Cross-Channel Prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–15 July 2017.
52. Li, D.; Hung, W.C.; Huang, J.B.; Wang, S.; Ahuja, N.; Yang, M.H. Unsupervised Visual Representation Learning by Graph-Based Consistent Constraints. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 678–694.
53. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep Clustering for Unsupervised Learning of Visual Features. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
54. Hoffer, E.; Hubara, I.; Ailon, N. Deep unsupervised learning through spatial contrasting. *arXiv* **2016**. https://doi.org/10.48550/arXiv.1610.00243.
55. Bojanowski, P.; Joulin, A. Unsupervised Learning by Predicting Noise. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Precup, D., Teh, Y.W., Eds.; Volume 70, pp. 517–526.
56. Li, Y.; Paluri, M.; Rehg, J.M.; Dollar, P. Unsupervised Learning of Edges. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, 26 June - 1 July 2016.
57. Purushwalkam, S.; Gupta, A. Pose from Action: Unsupervised Learning of Pose Features based on Motion. *arXiv* **2016**. https://doi.org/10.48550/arXiv.1609.05420.
58. Mahendran, A.; Thewlis, J.; Vedaldi, A. Cross Pixel Optical Flow Similarity for Self-Supervised Learning. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1807.05636.
59. Sayed, N.; Brattoli, B.; Ommer, B. Cross and Learn: Cross-Modal Self-Supervision. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1811.03879.
60. Korbar, B.; Tran, D.; Torresani, L. Cooperative Learning of Audio and Video Models from Self-Supervised Synchronization. In Proceedings of the 32nd International Conference on Neural Information Processing Systems—NIPS'18, Montréal Canada, 2–8 December 2018; Curran Associates Inc.: Red Hook, NY, USA, 2018; pp. 7774–7785.
61. Owens, A.; Efros, A.A. Audio-Visual Scene Analysis with Self-Supervised Multisensory Features. *arXiv* **2018**. https://doi.org10.48550/arXiv.1804.03641.
62. Kim, D.; Cho, D.; Kweon, I.S. Self-Supervised Video Representation Learning with Space-Time Cubic Puzzles. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence—AAAI'19/IAAI'19/EAAI'19, Honolulu, Hawaii, USA, 27 January–1 February 2019. https://doi.org/10.1609/aaai.v33i01.33018545.
63. Jing, L.; Yang, X.; Liu, J.; Tian, Y. Self-Supervised Spatiotemporal Feature Learning via Video Rotation Prediction. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1811.11387.
64. Fernando, B.; Bilen, H.; Gavves, E.; Gould, S. Self-Supervised Video Representation Learning with Odd-One-Out Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, Hawaii, USA, 22–25 July 2017; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 5729–5738. https://doi.org/10.1109/CVPR.2017.607.
65. Ren, Z.; Lee, Y. Cross-Domain Self-Supervised Multi-task Feature Learning Using Synthetic Imagery. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, Utah, USA, 18–22 June 2018; IEEE Computer Society: Los Alamitos, CA, USA, 2018; pp. 762–771. https://doi.org/10.1109/CVPR.2018.00086.

66. Wang, X.; He, K.; Gupta, A. Transitive Invariance for Self-Supervised Visual Representation Learning. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 1338–1347. https://doi.org/10.1109/ICCV.2017.149.
67. Doersch, C.; Zisserman, A. Multi-task Self-Supervised Visual Learning. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 2070–2079. https://doi.org/10.1109/ICCV.2017.226.
68. Mundhenk, T.; Ho, D.; Chen, B.Y. Improvements to Context Based Self-Supervised Learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; IEEE Computer Society: Los Alamitos, CA, USA, 2018; pp. 9339–9348. https://doi.org/10.1109/CVPR.2018.00973.
69. Noroozi, M.; Vinjimoor, A.; Favaro, P.; Pirsiavash, H. Boosting Self-Supervised Learning via Knowledge Transfer. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018; IEEE Computer Society: Los Alamitos, CA, USA, 2018; pp. 9359–9367. https://doi.org/10.1109/CVPR.2018.00975.
70. Büchler, U.; Brattoli, B.; Ommer, B. Improving Spatiotemporal Self-supervision by Deep Reinforcement Learning. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 797–814.
71. Liu, Y.; Jin, M.; Pan, S.; Zhou, C.; Zheng, Y.; Xia, F.; Yu, P. Graph Self-Supervised Learning A Survey. *IEEE Trans. Knowl. Data Eng.* **2022**. https://doi.org/10.1109/TKDE.2022.3172903.
72. Baevski, A.; Hsu, W.N.; Xu, Q.; Babu, A.; Gu, J.; Auli, M. data2vec A General Framework for Self-supervised Learning in Speech, Vision and Language. *arXiv* **2022**. https://doi.org/10.48550/arXiv.2202.03555.
73. Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; Tang, J. Self-supervised Learning Generative or Contrastive. *IEEE Trans. Knowl. Data Eng.* **2021**, *35*, 857–876. https://doi.org/10.1109/TKDE.2021.3090866.
74. Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste Self-Supervised Learning for Anomaly Detection and Localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (Online Conference), 19–25 June 2021; pp. 9664–9674.
75. Xie, Y.; Xu, Z.; Zhang, J.; Wang, Z.; Ji, S. Self-Supervised Learning of Graph Neural Networks A Unified Review. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**. https://doi.org/10.1109/TPAMI.2022.3170559.
76. Akbari, H.; Yuan, L.; Qian, R.; Chuang, W.H.; Chang, S.F.; Cui, Y.; Gong, B. VATT Transformers for Multimodal Self-Supervised Learning from Raw Video, Audio and Text. In Proceedings of the Advances in Neural Information Processing Systems, (Online Conference), 6–14 December 2021; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA 2021; Volume 34, pp. 24206–24221.
77. Kataoka, H.; Okayasu, K.; Matsumoto, A.; Yamagata, E.; Yamada, R.; Inoue, N.; Nakamura, A.; Satoh, Y. Pre-training without Natural Images. In Proceedings of the Asian Conference on Computer Vision (ACCV), (Online Conference), 30 November–4 December 2020.
78. Mahajan, D.; Girshick, R.; Ramanathan, V.; He, K.; Paluri, M.; Li, Y.; Bharambe, A.; van der Maaten, L. Exploring the Limits of Weakly Supervised Pretraining. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 185–201.
79. Li, W.; Wang, L.; Li, W.; Agustsson, E.; Van Gool, L. WebVision Database: Visual Learning and Understanding from Web Data. *arXiv* **2017**. https://doi.org/10.48550/arXiv.1708.02862.
80. Schmarje, L.; Santarossa, M.; Schröder, S.M.; Koch, R. A Survey on Semi-, Self- and Unsupervised Learning for Image Classification. *IEEE Access* **2021**, *9*, 82146–82168. https://doi.org/10.1109/ACCESS.2021.3084358.
81. Song, X.; Yang, H. A Survey of Unsupervised Learning in Medical Image Registration. *Int. J. Health Syst. Transl. Med. (IJHSTM)* **2022**, *2*.
82. Abukmeil, M.; Ferrari, S.; Genovese, A.; Piuri, V.; Scotti, F. A Survey of Unsupervised Generative Models for Exploratory Data Analysis and Representation Learning. *ACM Comput. Surv.* **2021**, *54*, 3450963. https://doi.org/10.1145/3450963.
83. Liu, T.; Yu, H.; Blair, R.H. Stability estimation for unsupervised clustering: A Review. *WIREs Comput. Stat.* **2022**, *14*, e1575. https://doi.org/10.1002/wics.1575.
84. Aoun, M.; Salloum, R.; Dfouni, A.; Sleilaty, G.; Chelala, D. A formula predicting the effective dose of febuxostat in chronic kidney disease patients with asymptomatic hyperuricemia based on a retrospective study and a validation cohort. *Clin. Nephrol.* **2020**, *94*, 61.
85. Zhang, R.; Isola, P.; Efros, A.A. Colorful Image Colorization. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 649–666.
86. Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context Encoders: Feature Learning by Inpainting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, USA, 26 June–1 July 2016.
87. Noroozi, M.; Favaro, P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 8–16 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 69–84. https://doi.org/10.1007/978-3-319-46466-4_5.

88. Xie, S.; Girshick, R.; Dollar, P.; Tu, Z.; He, K. Aggregated Residual Transformations for Deep Neural Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 22–25 July 2017; IEEE Computer Society: Los Alamitos, CA, USA, 2017; pp. 5987–5995. https://doi.org/10.1109/CVPR.2017.634.
89. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**. https://doi.org/10.48550/arXiv.1704.04861.
90. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.
91. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.
92. Noroozi, M.; Pirsiavash, H.; Favaro, P. Representation Learning by Learning to Count. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5899–5907. https://doi.org/10.1109/ICCV.2017.628.
93. Chen, T.; Kornblith, S.; Norouzi, M.; Hinton, G. A Simple Framework for Contrastive Learning of Visual Representations. In Proceedings of the 37th International Conference on Machine Learning, (Online Conference), 6–10 June 2020; III, H.D., Singh, A., Eds.; Volume 119, pp. 1597–1607.
94. He, K.; Fan, H.; Wu, Y.; Xie, S.; Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), (Online Conference), 14–19 June 2020; pp. 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975.
95. Mandelbrot, B.B.; Wheeler, J.A. The Fractal Geometry of Nature. *Am. J. Phys.* **1983**, *51*, 286–287, https://doi.org/10.1119/1.13295.
96. Landini, G.; Murray, P.I.; Misson, G.P. Local connected fractal dimensions and lacunarity analyses of 60 degrees fluorescein angiograms. *Investig. Ophthalmol. Vis. Sci.* **1995**, *36*, 2749–2755. Available online: https://arvojournals.org/arvo/content_public/journal/iovs/933408/2749.pdf (accessed on 14 December 2022).
97. Smith, T.; Lange, G.; Marks, W. Fractal methods and results in cellular morphology - dimensions, lacunarity and multifractals. *J. Neurosci. Methods* **1996**, *69*, 123–136. https://doi.org/10.1016/S0165-0270(96)00080-5.
98. Barnsley, M.F. *Fractals Everywhere*; Academic Press: Cambridge, MA, USA, 2014.
99. Monro, D.; Dudbridge, F. Rendering algorithms for deterministic fractals. *IEEE Comput. Graph. Appl.* **1995**, *15*, 32–41. https://doi.org/10.1109/38.364961.
100. Chen, Y.Q.; Bi, G. 3-D IFS fractals as real-time graphics model. *Comput. Graph.* **1997**, *21*, 367–370. https://doi.org/10.1016/S0097-8493(97)00014-9.
101. Pentland, A.P. Fractal-Based Description of Natural Scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *PAMI-6*, 661–674. https://doi.org/10.1109/TPAMI.1984.4767591.
102. Varma, M.; Garg, R. Locally Invariant Fractal Features for Statistical Texture Classification. In Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007; pp. 1–8. https://doi.org/10.1109/ICCV.2007.4408876.
103. Xu, Y.; Ji, H.; Fermüller, C. Viewpoint Invariant Texture Description Using Fractal Analysis. *Int. J. Comput. Vision* **2009**, *83*, 85–100.
104. Larsson, G.; Maire, M.; Shakhnarovich, G. FractalNet: Ultra-Deep Neural Networks without Residuals. In Proceedings of the 5th International Conference on Learning Representations—ICLR 2017, Toulon, France, 24–26 April 2017.
105. Falconer, K. *Fractal Geometry: Mathematical Foundations and Applications*; Wiley & Sons Ltd.: Hoboken, NJ, USA, 2004.
106. Farin, G. *Curves and Surfaces for Computer-Aided Geometric Design: A Practical Guide*; Academic Press: Cambridge, MA, USA, 1993.
107. Perlin, K.; Improving Noise *ACM Trans. Graph.* Association for Computing Machinery: New York, NY, USA, **2002**, *21*, 681–682. https://doi.org/10.1145/566654.566636.
108. Yamada, R.; Takahashi, R.; Suzuki, R.; Nakamura, A.; Yoshiyasu, Y.; Sagawa, R.; Kataoka, H. MV-FractalDB: Formula-driven Supervised Learning for Multi-view Image Recognition. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 2076–2083. https://doi.org/10.1109/IROS51168.2021.9635946.
109. Xie, J.; Zheng, Z.; Gao, R.; Wang, W.; Zhu, S.C.; Wu, Y.N. Learning Descriptor Networks for 3D Shape Synthesis and Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, t, USA, 18–22 June 2018.
110. Kanezaki, A.; Matsushita, Y.; Nishida, Y. RotationNet for Joint Object Categorization and Unsupervised Pose Estimation from Multi-View Images. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)* **2021**, *43*, 269–283. https://doi.org/10.1109/TPAMI.2019.2922640.
111. Kanezaki, A.; Matsushita, Y.; Nishida, Y. RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
112. Kataoka, H.; Matsumoto, A.; Yamada, R.; Satoh, Y.; Yamagata, E.; Inoue, N. Formula-driven Supervised Learning with Recursive Tiling Patterns. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), (Online Conference), 11–17 October 2021; pp. 4081–4088. https://doi.org/10.1109/ICCVW54120.2021.00455.
113. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. https://doi.org/10.1109/5.726791.

114. Li, Z.; Liu, F.; Yang, W.; Peng, S.; Zhou, J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, 1–21. https://doi.org/10.1109/TNNLS.2021.3084827.
115. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**. https://doi.org/10.48550/arXiv.2010.11929.
116. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jegou, H. Training data-efficient image transformers & distillation through attention. In Proceedings of the 38th International Conference on Machine Learning, (Online Conference), 18–24 July 2021; Meila, M., Zhang, T., Eds.; Volume 139, pp. 10347–10357.
117. Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 Million Image Database for Scene Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009.
118. Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. Master's thesis, University of Tront, Toronto, ON, Canada, 2009. Available online: https://www.cs.utoronto.ca/~kriz/learning-features-2009-TR.pdf (accessed on 8 April 2009).
119. Everingham, M.; Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The Pascal Visual Object Classes (VOC) Challenge. *Int. J. Comput. Vision* **2010**, *88*, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
120. Lake, B.M.; Salakhutdinov, R.; Tenenbaum, J.B. The Omniglot challenge: a 3-year progress report. *Curr. Opin. Behav. Sci.* **2019**, *29*, 97–104. https://doi.org/10.1016/j.cobeha.2019.04.007.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.