SynFine: Boosting Image Segmentation Accuracy Through Synthetic Data Generation and Surgical Fine-Tuning

Mehdi Mounsif[a], Yassine Motie[b], Mohamed Benabdelkrim[c] and Florent Brondolo[d]
Akkodis Research, France

Keywords: Computer Vision, Transfer Learning, Surgical Fine-Tuning, Synthetic Data Generation, Carbon Capture and Storage.

Abstract: Carbon Capture and Storage (CCS) has increasingly been suggested as one of the many ways to reduce CO_2 concentration in the atmosphere, hence tackling climate change and its consequences. As CCS involves robust modelling of physico-chemical mechanisms in geological formations, it benefits from CT-scans and accurate segmentation of rock core samples. Nevertheless, identifying precisely the components of a rock formation can prove challenging and could benefit from modern segmentation approaches, such as U-Net. In this context, this work introduces SynFine, a framework that relies on synthetic data generation and surgical fine-tuning to boost the performance of a model on a target data distribution with a limited number of examples. Specifically, after a pre-training phase on a source dataset, the SynFine approach identifies and fine-tunes the most responsive layers regarding the distribution shift. Our experiments show that, beyond an advantageous final performance, SynFine enables a strong reduction of the number of real-world labelled pairs for a given level of performance.

1 INTRODUCTION

In the recent years, many studies have highlighted the strong correlation between human activity and global warming, with CO_2 emissions being a major contributor to these dynamics (IPCC, 2022; Kramer et al., 2021). As the highly likely environmental modifications resulting from climate will have significant impacts on current societies and modern organisations, multiple institutions and actors have been developing policies, tools and methods (Kristjánsdóttir and Kristjánsdóttir, 2021) to try and reduce the anthropogenic effects on the greenhouse gases (Ashworth et al., 2010; Huaman and Jun, 2014; Wennersten et al., 2015) trapping the sun's radiation within the atmosphere. For instance, CCS (Carbon Capture and Storage) technologies intend on capturing carbon from emitters, and injecting its liquid form into deep geological formations. While multiple geographic areas are already known for their significant storage potential, important efforts have been invested in estimating the potential volume that can be cached in porous rocks.

For this purpose, understanding the elaborate physical, chemical and mechanical mechanisms involved during CO_2 flooding is of pivotal importance. It is a complex task that requires accurate modelling of the rock properties and is often based on the analysis and processing of CT-scanned volumes of rock samples. In this context, image segmentation, a task that consists in assigning a label to every pixel in an image such that pixels belonging to a similar class share certain characteristics, such as identifying all the rock pores of a CT-volume, is a commonly used approach to estimate rock porosity and adequately guide the modelling phase. While crucial, segmentation is, however, mostly done by hand and/or via histogram-based readings, involving user bias and errors, further leading to modelling inaccuracies (Saraf and Bera, 2021).

In this view, the remarkable technical progresses in computer vision, in particular through the increasing usage of deep neural networks, has enabled the scientific community to tackle tasks implying various modalities and highly diverse environments. Nevertheless, one of the drawbacks of these approaches is the significant amount of labelled data required to train supervised models and ensure their robustness. This is partially offset by techniques such as transfer learning which, beyond offering a significant training

[a] https://orcid.org/0000-0002-2763-3890
[b] https://orcid.org/0000-0001-6191-6547
[c] https://orcid.org/0000-0002-0978-3000
[d] https://orcid.org/0000-0002-9168-8474

Mounsif, M., Motie, Y., Benabdelkrim, M. and Brondolo, F.
SynFine: Boosting Image Segmentation Accuracy Through Synthetic Data Generation and Surgical Fine-Tuning.
DOI: 10.5220/0011848300003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 565-573
ISBN: 978-989-758-626-2; ISSN: 2184-4313

time reduction, enables to reuse features learned from an initial training task in other domains, making it a highly relevant strategy in the scope of industrial usage, due to potentially complex data gathering and labelling processes.

In this context, this paper introduces SynFine, a framework that enables the reduction of real-world labelled data needed for a given level of performance through the usage of synthetic data and surgical fine-tuning, a branch of transfer learning that focuses on specific parameters and challenges the common practice of training only the last layer. In particular, beyond shedding light on the strong synergy that can exist between transfer learning and synthetic data generation in an industrial scope, the contributions of this work are the following:

- Validating the surgical fine-tuning of a transfer learning approach for segmentation tasks on a dataset composed of corrupted images
- Providing a domain-driven example of procedural data generation
- Experimentally evaluating the SynFine method relevance regarding synthetic data generation and empirically proving that significant benefits, in particular labelling cost reduction, can be obtained in this setting

2 RELATED WORKS

Accurate modelling of rock physics is a promising and dynamic research direction (Ibrahim et al., 2021) that could potentially lead to significant workflow improvements in the geo-science industry and, in particular, CCS applications (Saraf and Bera, 2021). However, while notable progresses have been recently presented (Wang et al., 2018) in this particularly complex field, it is still typically approached using traditional statistical methods and could arguably benefit from the use of contemporary deep learning ideas.

Surveying modern AI technical landscape, the remarkable results and the highly diverse modalities tackled by models are undeniable (Devlin et al., 2018; Nichol et al., 2021) and hint at the possibilities of these approaches to extend to a wide array of domains, among which rock-physics modelling. One of the main limitations of learning-based models is access to qualitative and labelled data in sufficient volume, which can be challenging and incur heavy costs in practice (Everingham et al., 2015; Wu et al., 2021; Christiano et al., 2017).

Given these constraints, transfer learning (Oquab et al., 2014) has been one of the most popular strategies employed to train more robust models and reduce overfitting on modestly sized dataset. While this paradigm has been used through multiple domains (Razavian et al., 2014), it relies essentially on an unchanged workflow to adapt pre-trained features to target distributions. Specifically, multiple works (Kirkpatrick et al., 2017; Lee et al., 2019) demonstrate advantageous performances when the classifying layers of a model are fine-tuned while the rest of the parameters are frozen (Sener et al., 2016; Kirichenko et al., 2022). These strategies implement straightforward mechanisms to prevent the loss of information and have been extended with softer methods that involve weight regularization and network pruning (Myung et al., 2022). More recently, strong focus has been directed towards few-shots adaptation (Shen et al., 2021), general robustness (Andreassen et al., 2021) and model adaptation (Varsavsky et al., 2020), such as (Lee et al., 2022) on which this work heavily builds and that introduces the idea of specific layer responsiveness given the shift between the initial and target distribution.

The SynFine method presented in this work suggests that this fine-tuning paradigm can be particularly relevant in the context of synthetic data generation (Nikolenko, 2021). Indeed, as opposed to currently available manually-labelled datasets, which may lack support for fine-grained features and mostly implement masks as rough patches (Lin et al., 2014) due to the minutiae required, it is possible to generate pixel-perfect image-mask pairs through simulation and/or 3D modelling (Khan et al., 2019; Qiu et al., 2017). This accuracy has encouraged many works to rely on 3D pipelines as in (Johnson-Roberson et al., 2016), where the authors show that using a relatable synthetic dataset several orders of magnitudes more voluminous than its real counterpart enabled their model to reach a higher score on the real-world validation set. While this underlines the relevance of pre-training on synthetic dataset, the work proposed in this report suggests that potentially superior benefits can be reached through the introduction of specific fine-tuning strategies.

3 METHOD

As mentioned, the SynFine framework is heavily geared towards industrial considerations and, as such, suggests that the combined usage of surgical fine-tuning and synthetic datasets when real-world data is scarce or costly to produce is a relevant solution. In this view, this section first presents the general SynFine pipeline and explains the stages required by this method. Then, it provides detail regarding the data generation process which may present some particularities depending on the application context. Finally, it introduces the RGN (Ratio Norm Gradient) metric that can provide guidance regarding layer sensibility to fine-tuning and overall contribution to the model performance in the downstream task.

3.1 The SynFine Pipeline

The SynFine framework is inspired by the increasing interest in synthetic data generation approaches. While the remarkable recent improvements in this field can not be understated, it is unlikely that simulations will be able to replicate exactly the complex and highly non-linear processes and mechanisms underlying the physical world.

However, if real-world data were to be considered as a corrupted version of data generated through procedural functions, then it becomes possible to frame this configuration within the surgical fine-tuning paradigm. Specifically, following the ideas of layer sensibilities to distribution shift in the scope of transfer learning introduced in (Lee et al., 2022), SynFine, as presented in Figure 1 proposes the following workflow:

1. Produce an abundant labelled synthetic dataset, relatable to the target real-world dataset
2. Pre-train a given model on the synthetic data collection
3. Identify the most suitable and sensitive layers regarding the distribution shift between the synthetic dataset and the target real-world dataset
4. Apply surgical fine-tuning to maximise the pre-training benefits and final model performance in a possibly limited data regime.

3.2 Data Generation

While the SynFine paradigm is completely generic, there exist practical considerations that may significantly influence the data generation stage of the proposed method. In particular, since this phase aims at generating labelled synthetic data, it is crucial that the processes setup in this regard provide a way to isolate the mechanisms that produce the different labels related to the target dataset. As the current SynFine approach focuses mostly on expert-knowledge and manually designed pipelines to create labelled variations of images, as demonstrated in Section 4.1, more theoretical and causality-driven methods, for instance relying on conditional generative modelling, will be explored in further work.

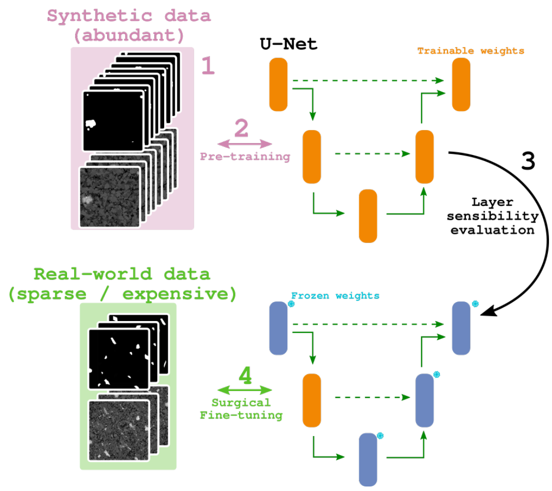

Figure 1: General view of the SynFine framework and the main workflow steps.

3.3 Surgical Fine-Tuning and Layer Sensibility Evaluation

As mentioned in Section 2, transfer learning has a paramount importance in modern deep learning approaches and applications. While the traditional workflow consists of an initial fine-tuning of the last layers and a progressive unfreezing of the earlier blocks, the authors of (Lee et al., 2022) show that, depending on the distribution shift between the source and target dataset, this strategy can prove sub-optimal and lead to lower performances than a surgical methodology that focuses the fine-tuning on the most suitable parameters.

Consequently, as suggested in Section 3.1, identifying layer sensibility is a crucial part of SynFine, in particular since the alternative approach would require to evaluate the final performance of each fine-tuning strategy, which is both inconvenient and inefficient. In this view, different metrics can be considered to compute layer sensibility to the distribution shift between the synthetic dataset and the real-world images. Specifically, the proposed pipeline has been evaluated under the RGN responsiveness strategy, which uses batches of data to estimate per-layer contribution to the error relative to the layer weights. Derived in Equation 1, with θ_i and g_i respectively being the parameters and the gradient of layer i.

$$\mathrm{RGN}(\theta_i) = \frac{g_i}{||\theta_i||} \tag{1}$$

4 EXPERIMENTS

In order to demonstrate the applicability and relevance of the SynFine framework, this section presents the multiple experiments setup in this view as well as the data created accordingly, central in this context. Specifically, following details regarding the processes used to generate mineral-inspired images and their segmentation masks, layer sensibility metrics are explored and the resulting performances for each fine-tuning strategies are analyzed to confirm the interest of the SynFine methodology.

For the experiments presented in this section, a regular 3 blocks U-Net model is implemented in PyTorch, with 467k parameters. Training phases rely on a Adam optimizer, with 2×10^{-3} learning rate for the initial pre-training phase and a 3×10^{-4} learning rate for the fine-tuning. Images and masks are downsampled to 64×64 patches and a batch size of 32 is used. Two thousand synthetic images were generated, which is one order of magnitude more than the real-world dataset. Finally, training and fine-tuning phases are both scheduled for 10 epochs.

4.1 Labelled Data Generation

While intrinsically generic, as mentioned in Section 3.2, some practical aspects of the target domain can imply specific workflows regarding the data generation pipeline, as exemplified in the following paragraphs that detail the process of mineral-inspired labelled image generation.

Complex organic shapes observable in nature can often be broken down into a non-linear combination and superposition of simpler patterns. Accordingly, the processing pipeline designed to implement these principles, visible in Figure 2, involves tiling multiple basic shapes while applying per-shape stretches and random scaling, rotations and offsets. These transformations are applied to each basic unit and the resulting collections of patterns are then blended using a max-intensity strategy which, contrary to an additive strategy, prevents saturating the pixels as this seems to be a notable feature of the target class.

While this first function provides the image background, the target class pixels can be created by applying a threshold function to the histogram of a noise texture as a starting image, consequently generating large areas without information and specks of clearer pixels. The application of a gaussian bluring filters can contribute in diffusing these pixels in wider areas which can then go through a step function to provide the desired masks. As feeding seeds to the noise generator will result in different masks, this approach is a convenient way to produce a large dataset for rock segmentation.

4.2 Surgical Fine-Tuning Validation on Corrupted Dataset

The main ideas of surgical fine-tuning were initially introduced in the scope of classification tasks. In these cases, input-level perturbations, among other types of distribution shift, were shown to strongly affect native performance. In this context, the first step towards the validation of the SynFine framework was to evaluate the system behaviour in a segmentation task with similar input corruptions.

As such, in this early-stage validation phase, no synthetic data is required since it rather aims at evaluating surgical fine-tuning technique in a segmentation context. In practice, as shown in Figure 3, corrupted data is derived from the initial rock segmentation dataset through the application of various visual perturbations.

After the initial training phase on the real dataset, the RGN sensibility metric, introduced in Section 3.3, is computed in order to provide guidance to the surgical fine-tuning phase. In practice, two ways of clustering the model parameters are analyzed due to the U-Net architecture specificity:

- The sequential view where blocks are considered independently, that is, all the encoding blocks followed by the decoding ones
- The transversal view, relying on a horizontal clustering of blocks the encoding/decoding pairs. In this case, parameters are gathered depending on their depth in the model.

Figure 4 displays the sensibility of each parameter group for both clustering paradigms. As can be seen, in both cases, the first/higher U-Net block indicates the highest responsiveness, which is coherent with the expected results since the perturbations added focus on input-level features. Since multiple batches of data are used, the shaded area represent the min-max variations while the solid line indicates the mean layer responsiveness value.

Finally, the final fine-tuning results for different freezing strategies, visible in Figure 5, confirm that surgical fine-tuning:

- Significantly outperforms *naive* fine-tuning
- Strongly influences the final model performance based on the parameters added to the trainable set
- Is also relevant in a segmentation context, despite the introduction of transversal gradients paths, inherent to the U-Net architecture.

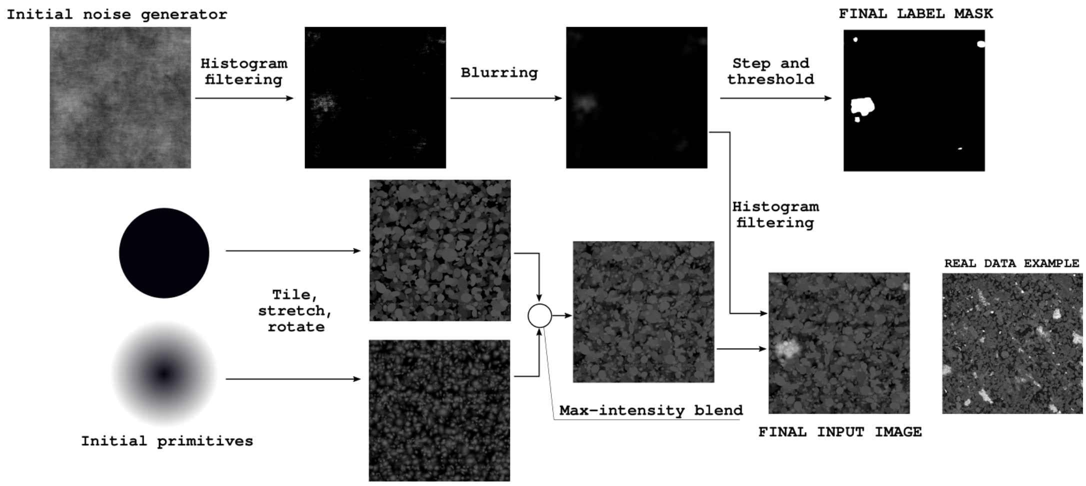

Figure 2: Mineal-inspired image generation process and reference image.

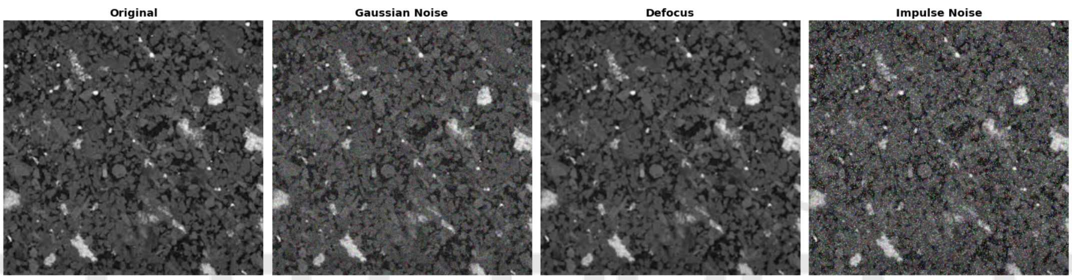

Figure 3: Instance of a real-world input image and multiple corruptions examples.

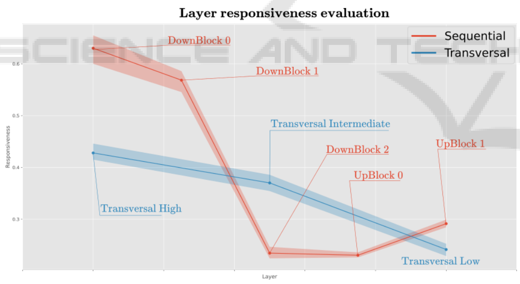

Figure 4: Layer responsiveness after initial training for RGN metrics on a corrupted version of the real-world dataset.

4.3 Transfer from Procedural Data to Real-World Images

This second experiment focuses on the main use case for the SynFine framework. Specifically, it considers the case where an abundant labelled synthetic dataset is available on which a segmentation model is initially trained. Then, using layer sensibility evaluation, the most responsive parameters regarding the distribution shift are identified and fine-tuned on the real-world rock segmentation dataset. In this view, Figure 6 shows the resulting accuracy levels for various fine-tuning strategies, along with the RGN sensibility per layer for both sequential and transversal view. While a range of final performances can be observed, every surgical fine-tuning strategy strongly outperforms the native model accuracy (that is, without fine-tuning) as well as the *naive* approach that consists in fine-tuning only the last layer, as summed up in Table 1.

It is however less trivial to interpret the RGN sensitivity measure, in particular in the Sequential configuration. Specifically, while the highest RGN scoring layer does indeed signal the most responsive parameters in both cases and follows the general shape of the surgical fine-tuning performance, the relation between RGN values and final scores of the DownBlock 1, 2 and UpBlock 0 is not entirely clear and will be further investigated in subsequent works.

Table 1: Surgical fine-tuning approaches performance increase with native model and *naive* fine-tuning.

Accuracy relative to	**DownBlock 0**	**Trans. High**
Native performance	26.6 ± 0.20	28.5 ± 0.27
Naive fine-tuning	35.9 ± 0.20	37.8 ± 0.28

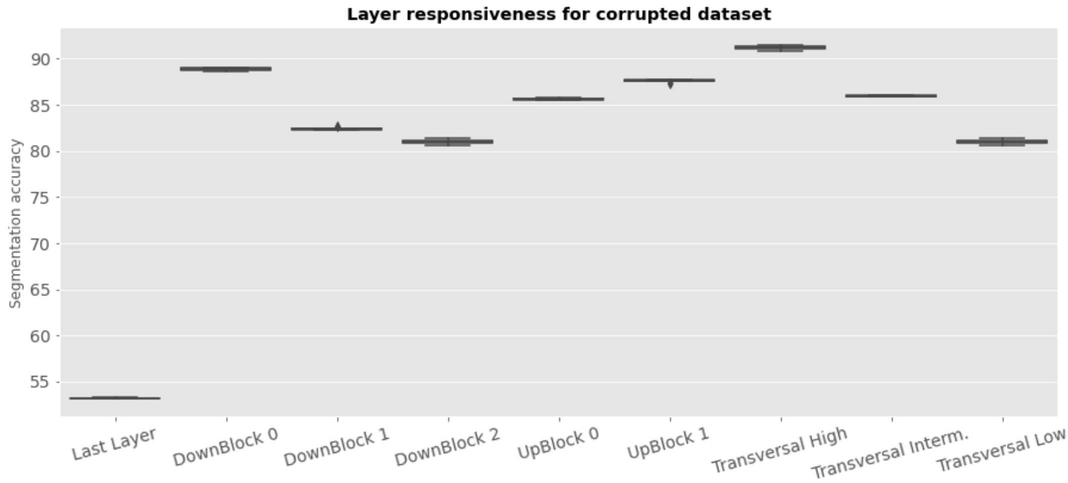

Figure 5: Accuracy comparison for different surgical fine-tuning target for both the sequential and horizontal clustering. From left to right, DownBlock 0-3, UpBlock 0-1 and the three depth levels of the transversal view.

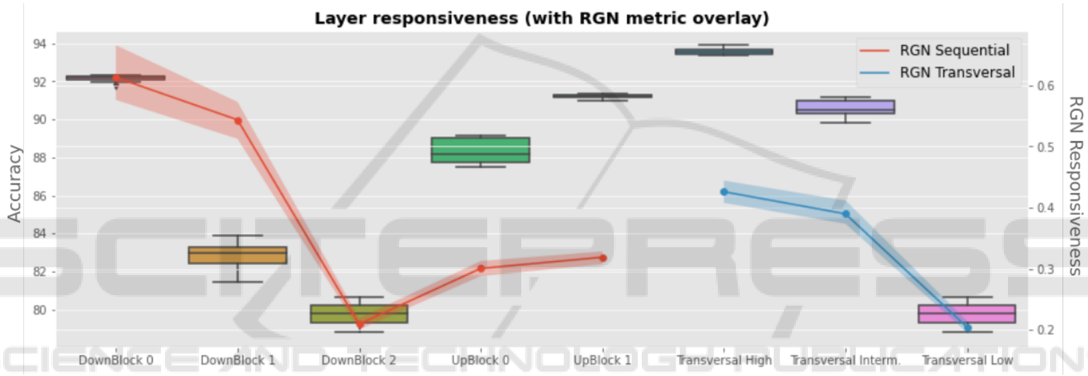

Figure 6: RGN layer responsiveness in a synthetic-to-real transfer for both parameters clustering paradigms.

4.4 Synthetic Advantages

The previous sections have provided experimental evidence of the relevance of surgical fine-tuning strategies for segmentation tasks and confirmed that correctly selecting which model parameters should be frozen during the optimization phase depending on the distribution shift can yield non-negligible increase in performance, contrary to *naive* approaches.

It is, however, also interesting to view the SynFine framework in the context of dataset creation. Indeed, experiments presented in Sections 4.2 and 4.3 focus on final performance for a given dataset size and volume. Nevertheless, in the rock segmentation configuration for instance, an operating company could very likely be aiming at minimizing the number of samples required to reach a given level of performance. In this paradigm, the per-sample efficiency is paramount and the SynFine approach can provide significant advantage, as can be seen in Figure 7. Specifically, three approaches are compared:

- A straightforward strategy, named Baseline and shown in violet, which does not rely on any form of transfer learning
- A *naive* transfer learning approach that fine-tunes only the last layer, labelled Traditional FT and displayed in blue
- Surgical FT, the responsiveness-driven surgical fine-tuning approach, in red.

As can be observed, the SynFine method displays a higher initial accuracy on the target dataset and shows that, in this configuration, labelled samples yield a more advantageous increase in performance than with other methods, thus underlining the strong interest of this approach in domains where data can be simulated but labelling is expensive. While the baseline performance gets close to the surgical fine-tuning when approaching the full dataset, it is likely that the gap could be more important if the images considered were more complex and will be explored in further works.

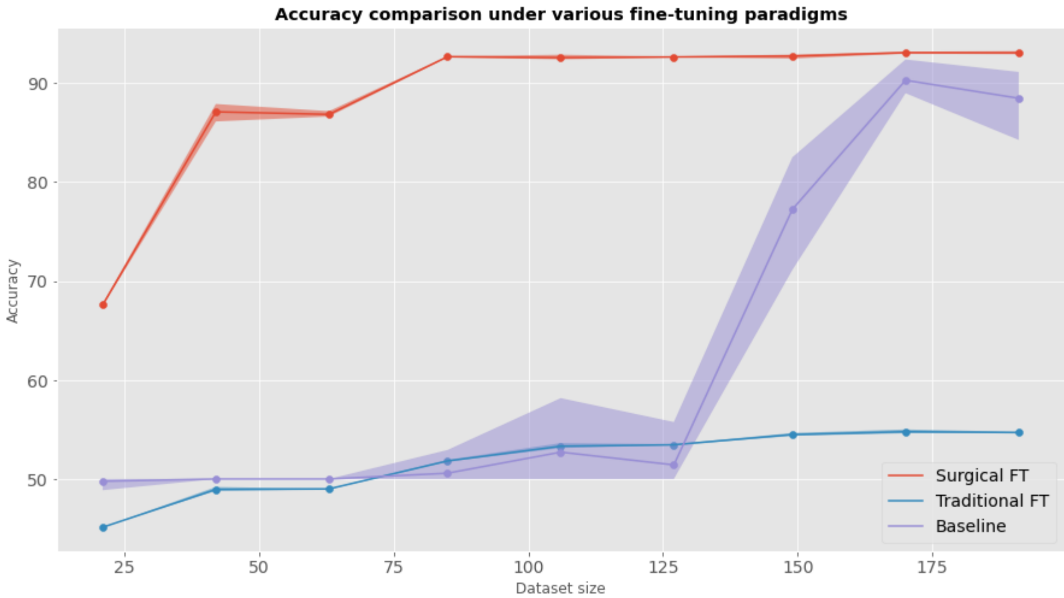

Figure 7: Segmentation accuracy comparison for multiple fine-tuning strategies against a baseline process for increasing dataset size (x-axis).

5 CONCLUSIONS

In this work, the SynFine framework, an empirical approach for maximizing transfer learning benefits for segmentation tasks in the scope of synthetic data generation is introduced. Through multiple experiments, displaying synthetic data generation strategies for real-world data distributions in different contexts, it has been shown that there exist significant advantages regarding performance when leveraging surgical fine-tuning. Additionally, this work provides local evidence that the SynFine framework can prove considerably more cost-effective than a *naive* approach, which is a central aspect in many real-world industries and applications. In particular, the improvements induced by the SynFine method could impulse significant advantages regarding CCS modelling and, *in fine*, deployment.

While having direct practical implications, the initial results presented in this work also open a vast array of perspectives. Among other considerations, it would be insightful to understand how to design synthetic data that triggers essentially intermediate U-Net blocks or levels, since most of the sensibility in our experiments was concentrated within the first block. This understanding could further be leveraged to help formulate a more theoretical compatibility measure between the target real-world dataset and the synthetically produced, which is lacking from the current approach.

Finally, regarding the synthetic data generation, while the presented strategy relies on expert knowledge to mass-produce labelled data, the increasing capabilities of generative models, in particular within the diffusion scope, and their ability to produce conditionally-driven images could also be considered to complement or completely replace the procedural approaches presented in this work and also provide more accessibility for domain of higher complexity.

ACKNOWLEDGEMENTS

This work has been sponsored by the Akkodis group.

REFERENCES

Andreassen, A., Bahri, Y., Neyshabur, B., and Roelofs, R. (2021). The evolution of out-of-distribution robustness throughout fine-tuning. *ArXiv*, abs/2106.15831.

Ashworth, P., Boughen, N., Mayhew, M., and Millar, F. (2010). From research to action: Now we have to move on ccs communication. *International Journal of Greenhouse Gas Control*, 4(2):426–433.

Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., and Amodei, D. (2017). Deep reinforcement learning from human preferences. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Infor-*

mation Processing Systems, volume 30. Curran Associates, Inc.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. cite arxiv:1810.04805Comment: 13 pages.

Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, 111(1):98–136.

Huaman, R. N. E. and Jun, T. X. (2014). Energy related co2 emissions and the progress on ccs projects: a review. *Renewable and Sustainable Energy Reviews*, 31:368–385.

Ibrahim, E., Jouini, M., Bouchaala, F., and Gomes, J. (2021). Simulation and validation of porosity and permeability of synthetic and real rock models using three-dimensional printing and digital rock physics. *ACS Omega*.

IPCC (2022). *Summary for Policymakers*. Cambridge University Press, Cambridge, UK. In Press.

Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S. N., and Vasudevan, R. (2016). Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? *CoRR*, abs/1610.01983.

Khan, S., Phan, B., Salay, R., and Czarnecki, K. (2019). Procsy: Procedural synthetic dataset generation towards influence factor studies of semantic segmentation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

Kirichenko, P., Izmailov, P., and Wilson, A. G. (2022). Last layer re-training is sufficient for robustness to spurious correlations.

Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., and Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526.

Kramer, R. J., He, H., Soden, B. J., Oreopoulos, L., Myhre, G., Forster, P. M., and Smith, C. J. (2021). Observational evidence of increasing global radiative forcing. *Geophysical Research Letters*, 48(7):e2020GL091585.

Kristjánsdóttir, H. and Kristjánsdóttir, S. (2021). Carbfix and sulfix in geothermal production, and the blue lagoon in iceland: Grindavík urban settlement, and volcanic activity. *Baltic Journal of Economic Studies*, 7(1):1–9.

Lee, J., Tang, R., and Lin, J. (2019). What would elsa do? freezing layers during transformer fine-tuning. *ArXiv*, abs/1911.03090.

Lee, Y., Chen, A. S., Tajwar, F., Kumar, A., Yao, H., Liang, P., and Finn, C. (2022). Surgical fine-tuning improves adaptation to distribution shifts.

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

Myung, S., Huh, I., Jang, W., Choe, J. M., Ryu, J., Kim, D., Kim, K.-E., and Jeong, C. (2022). Pac-net: A model pruning approach to inductive transfer learning. In *International Conference on Machine Learning*.

Nichol, A., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., Sutskever, I., and Chen, M. (2021). GLIDE: towards photorealistic image generation and editing with text-guided diffusion models. *CoRR*, abs/2112.10741.

Nikolenko, S. (2021). *Synthetic Data for Deep Learning*. Springer Optimization and Its Applications. Springer International Publishing.

Oquab, M., Bottou, L., Laptev, I., and Sivic, J. (2014). Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1717–1724.

Qiu, W., Zhong, F., Zhang, Y., Qiao, S., Xiao, Z., Kim, T. S., and Wang, Y. (2017). UnrealCV: Virtual worlds for computer vision. In *Proceedings of the 25th ACM International Conference on Multimedia*, page 1221–1224. Association for Computing Machinery.

Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). Cnn features off-the-shelf: an astounding baseline for recognition.

Saraf, S. and Bera, A. (2021). A review on pore-scale modeling and ct scan technique to characterize the trapped carbon dioxide in impermeable reservoir rocks during sequestration. *Renewable and Sustainable Energy Reviews*, 144:110986.

Sener, O., Song, H. O., Saxena, A., and Savarese, S. (2016). Learning transferrable representations for unsupervised domain adaptation. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Shen, Z., Liu, Z., Qin, J., Savvides, M., and Cheng, K.-T. (2021). Partial is better than all: Revisiting fine-tuning strategy for few-shot learning. *ArXiv*, abs/2102.03983.

Varsavsky, T., Orbes-Arteaga, M., Sudre, C. H., Graham, M. S., Nachev, P., and Cardoso, M. J. (2020). Test-time unsupervised domain adaptation. In Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D., and Joskowicz, L., editors, *Medical Image Computing and Computer Assisted Intervention*. Springer International Publishing.

Wang, B., Chen, Y., Lu, J., and Jin, W. (2018). A rock physics modelling algorithm for simulating the elastic parameters of shale using well logging data. *Scientific Reports*, 8.

Wennersten, R., Sun, Q., and Li, H. (2015). The future potential for carbon capture and storage in climate change mitigation–an overview from perspectives of technology, economy and risk. *Journal of Cleaner Production*, 103:724–736.

Wu, J., Ouyang, L., Ziegler, D. M., Stiennon, N., Lowe, R., Leike, J., and Christiano, P. F. (2021). Recursively summarizing books with human feedback. *ArXiv*, abs/2109.10862.