
Использование пространственно-временных зависимостей для мониторинга аддитивного производства с применением глубокого обучения

A Preprint

Vladislav Soldatov
worklikewalkietalkie@gmail.com

Dmitry Kropotov
dmitry.kropotov@gmail.com

Abstract

Мониторинг процессов аддитивного производства в режиме реального времени необходим для обеспечения качества деталей и предотвращения дефектов, так как лазерное сплавление в порошковом слое – это динамический и сложный процесс. Однако существующие алгоритмы мониторинга в режиме реального времени часто зависят от конкретного случая и не устойчивы к изменениям параметров процесса или типов данных. В данной работе рассмотрен один из методов глубокого обучения, ранее использованных для решения этой проблемы, а также предложен новый подход к задаче. Кроме того, было проведено исследование вышеупомянутых алгоритмов и их сравнение на наборе данных NIST «Overhang Part X4».

Keywords Additive manufacturing · Laser Powder Bed Fusion · Online monitoring · Flaw detection · Porosity prediction · Spatial-temporal Modeling

1 Введение

Аддитивное производство (АП) – это быстро развивающаяся технология, которая потенциально способна произвести революцию в обрабатывающей промышленности Ko et al. [2015]. Технология АП позволяет создавать сложные, легкие и высокопроизводительные детали, которые было бы трудно или невозможно изготовить традиционными методами. Однако процессы АП сложны и могут быть подвержены дефектам. Использование аддитивного производства в промышленности в настоящее время ограничено из-за неизвестного заранее и ненадежного качества деталей Grasso and Colosimo [2017]. Это в значительной степени вызвано сложными взаимосвязями между параметрами процесса, такими как мощность лазера, скорость лазерного излучения и другими настройками машины. Поэтому важно разработать методы мониторинга процессов АП в режиме реального времени для обеспечения качества деталей.

Несмотря на возможность использования статистических Yang et al. [2022] методов и классических алгоритмов машинного обучения Yeung et al. [2020], одним из многообещающих подходов к мониторингу процессов АП в режиме реального времени является использование глубокого обучения (DL). В данной работе представлен новый подход к использованию DL для мониторинга процессов АП в режиме реального времени и проведено сравнение с одним из лучших алгоритмов для данной задачи на настоящий момент, ConvLSTMMAE Ko et al. [2022]. Как предлагаемый, так и существующий подходы используют пространственно-временную модель для изучения взаимосвязей между параметрами процесса и качеством детали. Пространственно-временная модель учитывает динамическую природу процессов АП. Это позволяет рассматриваемым подходам быть более точными, чем традиционные методы, которые учитывают только статические зависимости между параметрами процесса и качеством детали.

2 Обзор задачи

(Рассматриваются в главах 1-2)

Процессы АП сильно варьируются в зависимости от таких факторов, как свойства используемых материалов, качество платформы для сборки и квалификация оператора. Эта изменчивость затрудняет разработку моделей, которые могут точно предсказать поведение процессов АП при любых условиях. Например, свойства используемых материалов могут влиять на температуру плавления металла, которую необходимо учитывать при построении алгоритма, использующего физические признаки управляемого процесса.

Помимо этого, процессы АП очень динамичны и включают в себя множество взаимодействующих факторов. Например, всплески ванны лазерного расплава могут влиять на поток расплавленного металла, что, в свою очередь, может повлиять на форму и пористость изготавливаемой детали.

Таким образом, основная сложность исследований в сфере мониторинга процессов АП заключается в том, что существует большое количество разнообразных установок, отличающихся производственными возможностями, настройками и параметрами. В связи с этим довольно трудоёмким оказывается процесс стандартизации и обобщения условий, в которых метод мониторинга АП должен осуществлять работу. В результате, создание универсальной модели, которую можно было бы использовать для мониторинга всех машин АП, становится затруднено.

Другой проблемой является нехватка наборов данных, на основе которых можно обучать и тестировать модели DL. Наборы данных для мониторинга процессов АП обычно собираются с использованием систем с замкнутым контуром Freeman et al. [2020]. Системы с замкнутым контуром контролируют процесс АП в режиме реального времени и используют эту информацию для настройки параметров процесса. Это затрудняет сбор наборов данных, репрезентативных для всего спектра возможных условий процесса АП. Кроме того, организации, занимающиеся подобными исследованиями, проводят их на наборах закрытых данных для внутреннего пользования, собранных на конкретной установке, доступной соответствующей организации.

В данной работе основной фокус внимания был уделён направлению АП под названием «Лазерное сплавление в порошковом слое» (Laser Powder Bed Fusion или LPBF), а в качестве данных использовались изображения лазерных ванн расплава (см. Рис. 2). Для обучения моделей применялся набор данных «Overhang Part X4» Lane and Yeung [2020]. В связи с большим объёмом неразмеченных данных, задача состояла в обучении нейросетей представлениям идущих подряд во времени кадров ванн расплава. Модель должна выдавать в качестве ответа изображение прогнозируемого следующего после входных данных кадра. Это позволяет перейти от постановки задачи в терминах всевозможных используемых материалов и установок АП, к пространству последовательностей кадров ванн расплава, имеющих характерные изменения в форме при проявлении аномалий, идентичные для любых характеристик установок и материалов.

Мат постановка? Мат описание метода?

3 Условия экспериментов

Эксперимент.

3.1 Методология

В качестве подхода машинного обучения к задаче в вышеописанной постановке, была выбрана парадигма самообучения (self-supervised learning) на основе автокодировщиков. Модели данного типа могут обнаруживать сложные зависимости в данных, создавая содержательные представления неразмеченных объектов. Процесс создания модели для обнаружения аномальных объектов и последующего её применения состоит из следующих этапов:

1. Определить и выделить «нормальные» (не являющиеся аномальными) объекты выборки.
2. Обучить автокодировщик на выделенной подвыборке.
3. По валидационной выборке выбрать порог для значения функционала ошибки восстановления между предсказанным кадром и правильным ответом.
4. При превышении значения порога на функционал ошибки восстановления, объявлять кадры текущей последовательности аномальными.

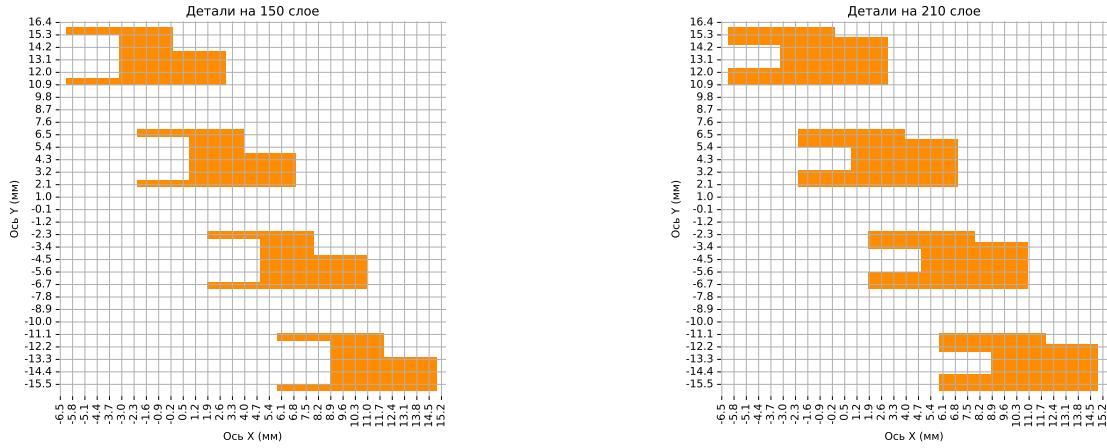
В качестве функционала ошибки восстановления была выбрана квадратичная ошибка между кадром из выборки и предсказанным кадром:

$$\text{SquaredError}(X, \hat{X}) := \sum_{i,j} (X_{ij} - \hat{X}_{ij})^2 = \|X - \hat{X}\|_F^2, \quad (1)$$

где X и \hat{X} – матрицы пиксельных представлений вышеуказанных кадров.

3.2 Обзор данных

В выбранном наборе данных представлены последовательные кадры лазерных ванн расплава, полученные при послойном наплавлении 4 одинаковых деталей, имеющих «нависания», то есть структурные компоненты такие, что на более поздних слоях плавления, новые части детали не имеют под собой опоры в виде уже наплавленного материала. Всего в процессе сбора данных запечатлено 250 последовательных слоёв печати по 20 . Каждая из 4 деталей имеет размер $5 \text{ mm} \times 9 \text{ mm} \times 5 \text{ mm}$ и выплавлена на пластине из кованого никелевого сплава (IN625). Направление плавления меняется на 90° от слоя к слою. Геометрия деталей на нескольких слоях отображена на Рис. 1.



(a) Взаимное расположение деталей на слое №150.

(b) Взаимное расположение деталей на слое №210.

Рис. 1: Геометрия деталей на слоях №150 и №210.

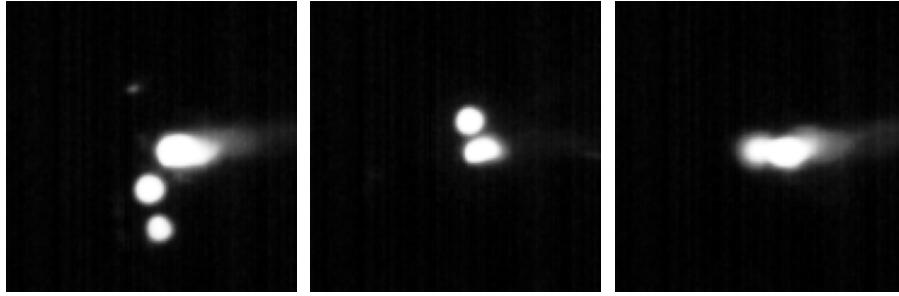
Изображения ванн расплава в наборе данных имеют размер 120×120 пикселей со значениями в диапазоне от 0 до 255. Каждый пиксель соответствует площади $8 \text{ мкм} \times 8 \text{ мкм}$ на лазерной установке.

3.3 Структуризация и предобработка данных

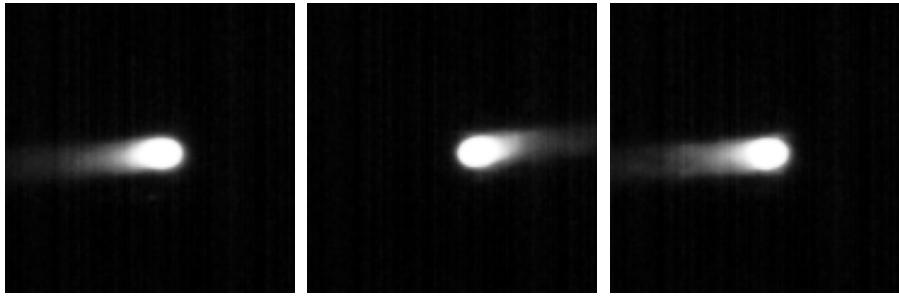
В связи с большим объёмом набора данных, было принято решение обучать модели на одном из 250 слоёв, и проводить валидацию на другом. Для обучения был выбран слой под номером 210, а для тестирования алгоритма – слой №150, аналогично Ko et al. [2022]. Первым шагом обработки выборки было полуавтоматическое определение некоторого количества аномальных изображений, согласно следующим правилам:

- Если на кадре присутствует несколько пятен соотносимого размера (это означает, что в данный момент произошёл всплеск, способный повлиять на качество детали)
- Если в последовательности подряд идущих кадров начинается резкое уменьшение/увеличение площади ванны расплава (это может означать, что произошла неполадка в работе установки или лазер начал двигаться по месту, где на детали остался дефект от одного из предыдущих проходов)

На Рис. 2а и 2б отображены соответственно аномальные и «нормальные» примеры ванн расплава, согласно описанному разделению.



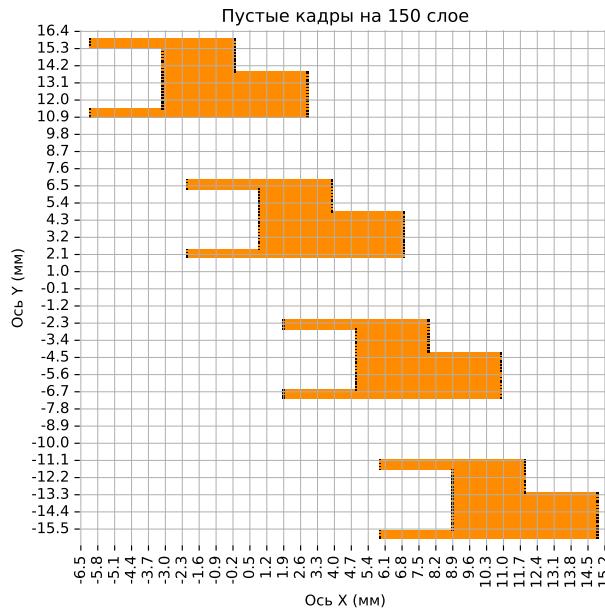
(a) Примеры аномальных кадров.



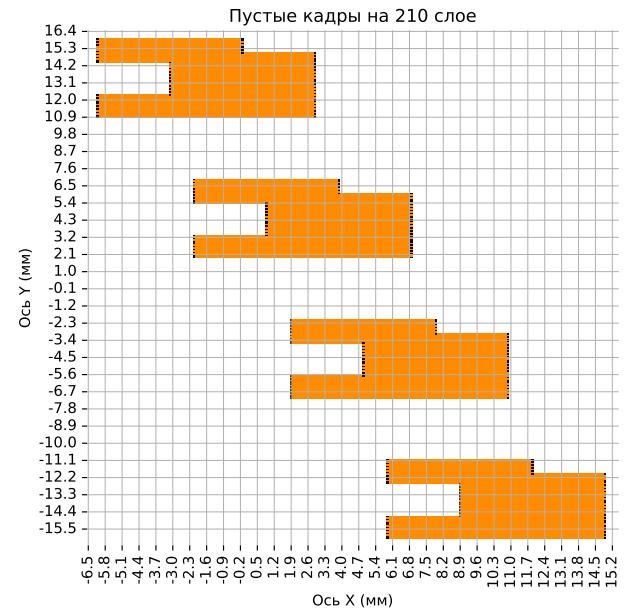
(b) Примеры «нормальных» кадров.

Рис. 2: Примеры изображений ванн расплава.

При подробном рассмотрении данных были обнаружены «пустые» кадры, на которых нет ванны расплава. Такие кадры встречаются на краях деталей (см. Рис. 3), в связи с тем, что лазер заканчивает проход в одну сторону, и начинает в обратную. Подобные изображения тоже были отмечены как неподходящие для обучения.



(a) «Пустые» кадры на слое №150.



(b) «Пустые» кадры на слое №210.

Рис. 3: Расположение «пустых» кадров на слоях №150 и №210, отмечены чёрным.

Все значения пикселей на изображениях были переведены в диапазон [0.0, 1.0]. На вход моделям DL подавались последовательности из нескольких (фиксированное в пределах одной модели значение) подряд идущих во времени кадров, а в качестве прогноза ожидался кадр, следующий за входом. В связи с тем, что в обучающей выборке была помечена неподходящими для обучения часть изображений, вышеуказанные последовательности формировались следующим образом: если «помеченный» кадр попадал в последовательность или следовал сразу за ней, то такая последовательность исключалась из тренировочного набора данных. Полученные таким образом выборки будем называть далее по номеру соответствующего слоя с приписанной буквой А (Пример: набор данных 150А).

В качестве аугментаций, для повышения обобщающей способности моделей, были использованы следующие приёмы: поворот всей последовательности кадров (включая следующий за ней предсказываемый кадр) на угол из диапазона $[-2\pi, 2\pi]$, и изменение масштаба в диапазоне [70%, 130%].

В дальнейшем для получения лучшей модели был использован следующий подход: производилось обучение алгоритма на полученных в результате применения вышеописанных шагов данных, а потом по порогу на функционал ошибки восстановления (1) из обучающей выборки аналогичным образом удалялись последовательности, определённые полученной моделью как аномальные. Полученные в результате этих действий данные будем обозначать по номеру слоя с приписанной буквой В (Пример: набор данных 210В). Затем новая модель обучалась на вновь очищенным от аномалий наборе данных.

4 Существующий подход

В качестве первого метода была рассмотрена модель ConvLSTMMAE Ko et al. [2022] – автокодировщика, основанного на архитектуре свёрточной LSTM Shi et al. [2015]. Такая архитектура позволяет успешно использовать последовательную структуру данных, являющихся при этом двумерными сигналами. Подробное описание архитектуры использовавшейся нейронной сети можно наблюдать в Приложении А. Обучение производилось с помощью оптимизатора AdamW Loshchilov and Hutter [2019] с темпом обучения равным 10^{-3} и с использованием функции потерь бинарной кросс-энтропии:

$$BCELoss(X, \hat{X}) := - \sum_{i,j} \left(X_{ij} \log \hat{X}_{ij} + (1 - X_{ij}) \log (1 - \hat{X}_{ij}) \right), \quad (2)$$

где X – матрица целевого кадра, а \hat{X} – предсказанного. Размер мини-батча был выбран равным 16, обучение проводилось на протяжении 20 эпох. Данные гиперпараметры обучения моделей были зафиксированы для всех алгоритмов на протяжении всего исследования.

4.1 Эксперименты

Согласно описанному выше подходу, сначала производилось обучение на выборке 210А с проведением валидации на наборе данных 150А. Для первой модели было зафиксировано значение числа кадров в последовательности равное 4. На Рис. 4а и 4б изображены примеры из соответственно тренировочного и валидационного наборов худших по значению функционала ошибки восстановления кадров.

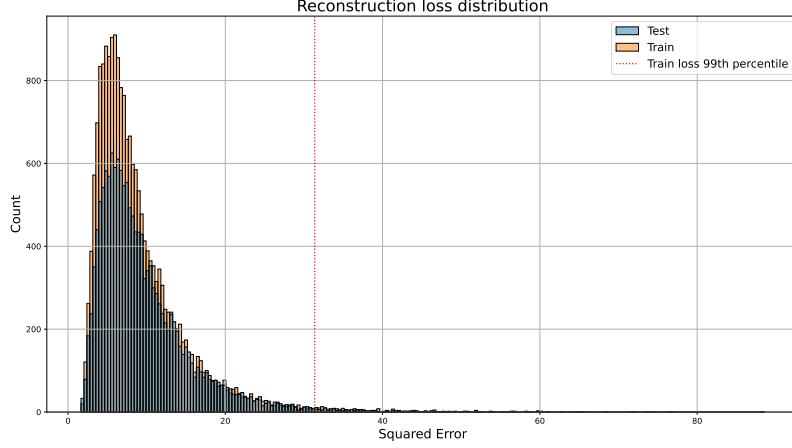
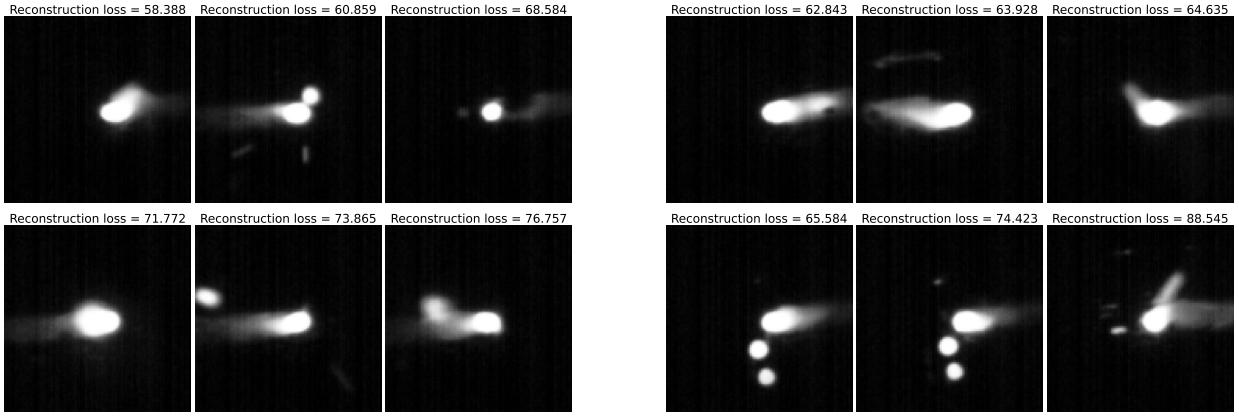


Рис. 5: Распределение значений функционала (1) на обучающей и валидационной выборке для модели ConvLSTMMAE, обученной на 210A при значении размера окна равном 4.



(a) Примеры худших по значению функционала ошибки восстановления кадров из тренировочного набора 210A.

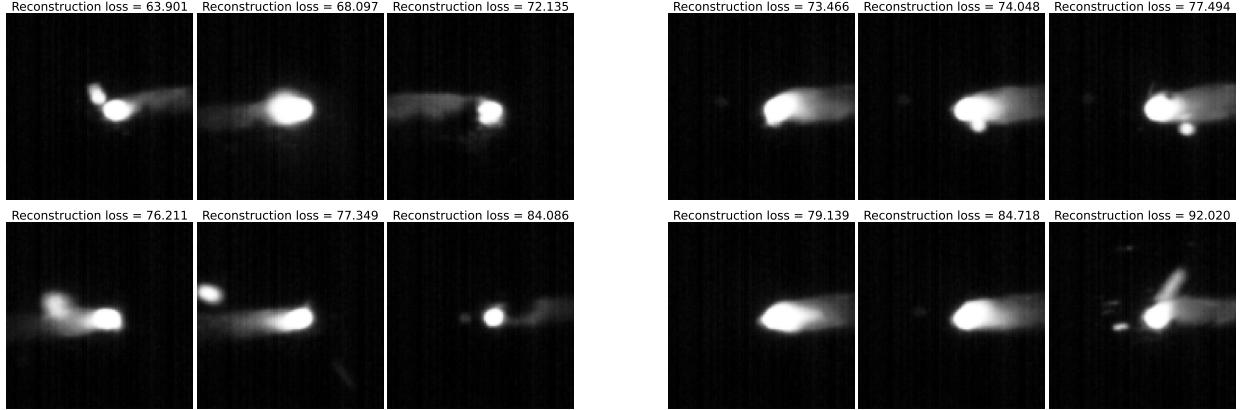
(b) Примеры худших по значению функционала ошибки восстановления кадров из валидационного набора 150A.

Рис. 4: Примеры найденных ConvLSTMMAE, обученной на 210A с размером окна 4, аномальных кадров. Над изображениями подписано значение функционала ошибки восстановления (1).

На Рис. 5 изображена плотность распределения значения функционала ошибки восстановления полученной модели на обучающей и на тестовой выборках, и отмечено значение 99-го перцентиля значения SE (1) на тренировочном наборе данных.

Видно, что плотности на обучающей и валидационной выборках мало отличаются по «форме», что говорит о присутствии генерализации модели и оправданности её применения для мониторинга в течение всего процесса плавления. При этом заметно, что распределение имеет «тяжёлый хвост», что может говорить о том, что модель имеет сравнительно высокую ошибку на объектах, не являющихся аномальными.

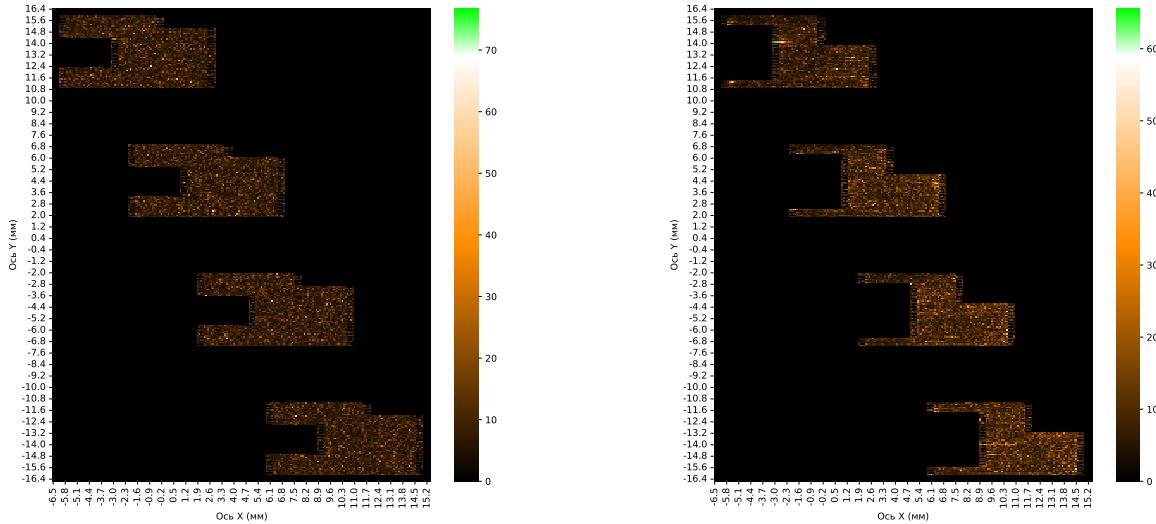
На Рис. 6 отображены значения функционала ошибки восстановления для различных координат положения лазера в момент запечатления ванны расплава. В связи с пустыми кадрами, появляющимися при изменении направления движения лазера на 180° , для таких последовательностей предсказания не были учтены во избежание смещения фокуса рассмотрения на места с заведомо большим значением ошибки. Координаты ванн расплава взяты с округлением в 0.1 мм, проведено усреднение значений функционала ошибки по координате. В связи со спецификой выбранного валидационного набора данных, на Рис. 6б можно заметить большие аномальные значений функционала ошибки, чем на Рис. 6а.



(a) Примеры худших по значению функционала ошибки восстановления кадров из тренировочного набора 210В.

(b) Примеры худших по значению функционала ошибки восстановления кадров из валидационного набора 150А.

Рис. 7: Примеры найденных ConvLSTMAE, обученной на 210В с размером окна 4, аномальных кадров. Над изображениями подписано значение функционала ошибки восстановления (1).



(a) Среднее по координате значение функционала ошибки восстановления для тренировочного набора 210А.

(b) Среднее по координате значение функционала ошибки восстановления для валидационного набора 150А.

Рис. 6: Распределение функционала (1) по координате для модели ConvLSTMAE, обученной на 210А с размером окна 4.

Такое представление позволяет сказать больше о возможных нарушениях производимого процесса печати. Например, на области, в которой в среднем происходит много аномалий, может оказаться дефективное наплавление, которое будет оказывать влияние на технологический процесс на протяжении нескольких последующих слоёв. В таком случае необходимо останавливать процесс печати для исправления ситуации или прекращения работы в случае неисправимой ошибки во избежание потерь времени и материалов на производство заведомо бракованной детали.

Для проведения второй части описанного метода обучения модели, по алгоритму из подраздела 3.3 из набора данных 210А были удалены 1000 объектов, являющихся худшими по значению функционала ошибки восстановления для модели ConvLSTMAE с размером окна 4.

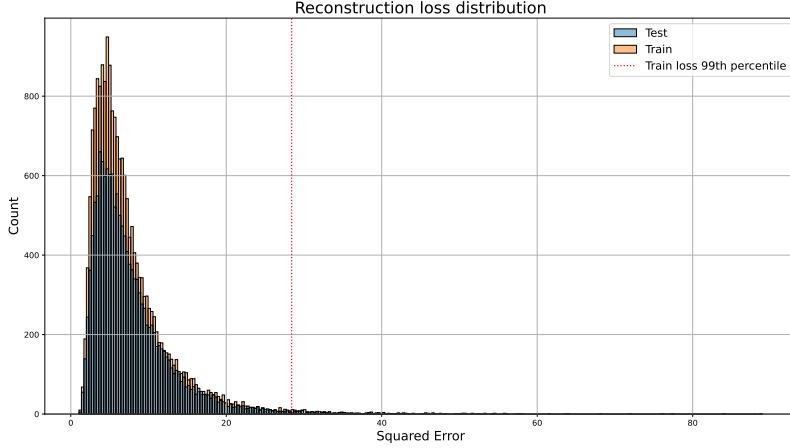


Рис. 8: Распределение значений функционала (1) на обучающей и валидационной выборке для модели ConvLSTMAE, обученной на 210В при значении размера окна равном 4.

На полученном с помощью модели ConvLSTMAE наборе данных 210В, была вновь обучена модель той же архитектуры. На Рис. 7 изображены найденные новой моделью аномальные кадры из наборов данных 210В и 150А. Можно отметить, что худшие по функционалу ошибки восстановления кадры из набора 210В визуально не столь очевидно представляют опасность для процесса плавления, чем удалённые в результате очистки выборки 210А (см. Рис. 4а), но всё так же являются результатами аномальных ситуаций (имеются в виду чёрные полости в ваннах расплава, свидетельствующие о мешающем лазеру наплавлении, а также сравнимые по площади с ванной всплески).

На Рис. 8 изображены распределения функционала ошибки восстановления итоговой модели на наборах данных 210А и 150А. Можно отметить, что в сравнении с Рис. 5, распределение ошибки на тестовой выборке стало более концентрированным, более «прижатым» к нулю, а значит и более похожим на распределение на обучающей выборке. Это свидетельствует о хорошей обобщающей способности получившегося алгоритма, а также о его приспособленности к применению к задаче поиска аномалий.

Помимо описанного выше исследования модели с размером окна, равным 4, были поставлены эксперименты по исследованию влияния иных значений данного гиперпараметра для модели ConvLSTMAE. В частности, были учтены значения в 2 и 6 кадров, составляющих последовательность, подающуюся на вход нейронной сети. Каждая из моделей была обучена на «очищенном» наборе данных 210В. На Рис. 9 и 10 можно наблюдать распределения функционала ошибки восстановления, получаемые с помощью таких моделей, а распределения ошибки (1) по координатной плоскости для всех методов отображены в Приложении В. Сравнение всех рассматриваемых в работе моделей представлено в разделе ??.

Список литературы

- Hyunwoong Ko, Seung Ki Moon, and Jihong Hwang. Design for additive manufacturing in customized products, 2015. URL <https://doi.org/10.1007/s12541-015-0305-9>.
- Marco Grasso and Bianca Maria Colosimo. Process defects and in situ monitoring methods in metal powder bed fusion: a review, 2017.
- Hui Yang, Siqi Zhang, Yan Lu, Paul Witherell, and Soundar Kumara. Spatiotemporal monitoring of melt-pool variations in metal-based additive manufacturing, 2022.
- Ho Yeung, Zhuo Yang, and Yan Lu. A meltpool prediction based scan strategy for powder bed fusion additive manufacturing, 2020.
- Hyunwoong Ko, Jaehyuk Kim, Yan Lu, Dongmin Shin, Zhuo Yang, and Yosep Oh. Spatial-temporal modeling using deep learning for real-time monitoring of additive manufacturing, 2022. URL <https://doi.org/10.1115/detc2022-91021>.
- Felicity S H B Freeman, Lova Chechik, and Iain Todd. Beat the machine (learning): metal additive manufacturing and closed loop control, 2020. URL <https://doi.org/10.1088/1361-6552/ab9957>.

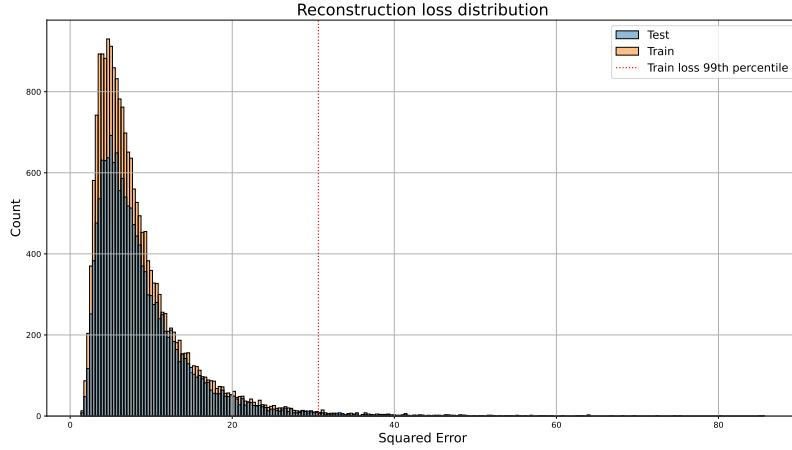


Рис. 9: Распределение значений функционала (1) на обучающей и валидационной выборке для модели ConvLSTMMAE, обученной на 210В при значении размера окна равном 2.

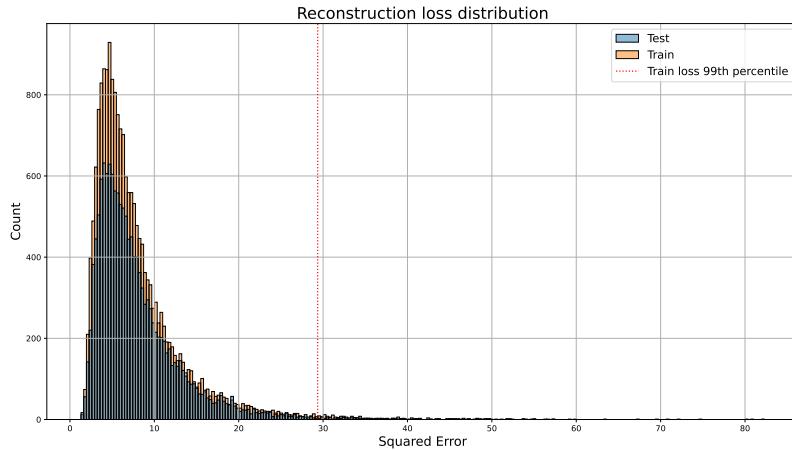


Рис. 10: Распределение значений функционала (1) на обучающей и валидационной выборке для модели ConvLSTMMAE, обученной на 210В при значении размера окна равном 6.

Brandon Lane and Ho Yeung. Process monitoring dataset from the additive manufacturing metrology testbed (ammt): Overhang part x4, 2020. URL <https://data.nist.gov/od/id/mds2-2233>.

Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai kin Wong, and Wang chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting, 2015.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

Yarin Gal and Zoubin Ghahramani. A theoretically grounded application of dropout in recurrent neural networks, 2016.

A Архитектуры нейронных сетей

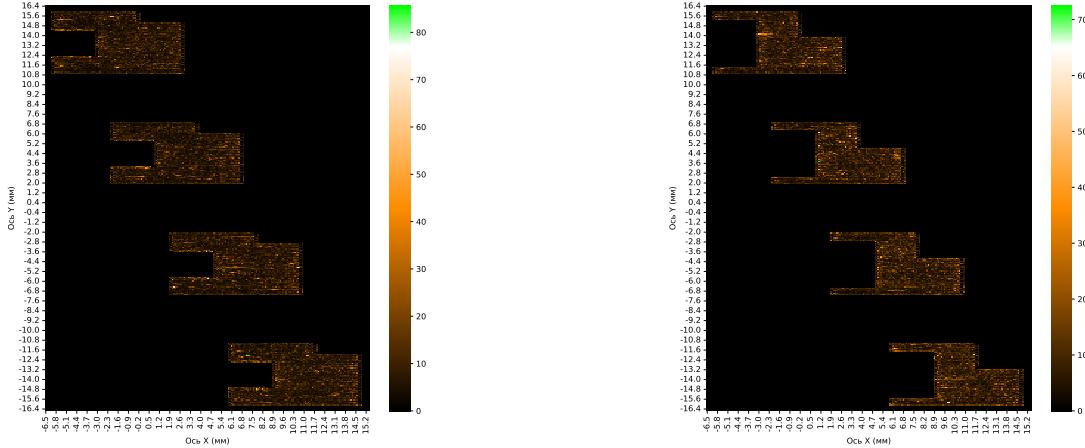
	Слой	Число каналов (вход/выход)	Размер фильтра	Шаг	Доп. действия	Активация
Кодировщик	Conv3D	1/64	$3 \times 3 \times 3$	1	-	ReLU
	Conv3D	64/64	$3 \times 3 \times 3$	1	Сохранить активацию 1	ReLU
	MaxPool3D	64/64	$1 \times 2 \times 2$	$1 \times 2 \times 2$	-	-
	Conv3D	64/128	$3 \times 3 \times 3$	1	-	ReLU
	Conv3D	128/128	$3 \times 3 \times 3$	1	Сохранить активацию 2	ReLU
	MaxPool3D	128/128	$1 \times 2 \times 2$	$1 \times 2 \times 2$	-	-
	Conv3D	128/256	$3 \times 3 \times 3$	1	-	ReLU
	Conv3D	256/256	$3 \times 3 \times 3$	1	-	ReLU
Декодировщик	MaxPool3D	256/256	$1 \times 2 \times 2$	$1 \times 2 \times 2$	-	-
	Interpolation	256/256	$1 \times 2 \times 2$	-	-	-
	Conv3D	256/128	$3 \times 3 \times 3$	1	Прибавить активацию 2	-
	Conv3D	128/128	$3 \times 3 \times 3$	1	-	ReLU
	Conv3D	128/128	$3 \times 3 \times 3$	1	-	ReLU
	Interpolation	128/128	$1 \times 2 \times 2$	-	-	-
	Conv3D	128/64	$3 \times 3 \times 3$	1	Прибавить активацию 1	-
	Conv3D	64/64	$3 \times 3 \times 3$	1	-	ReLU
Выход	Conv3D	64/1	$1 \times 1 \times 1$	1	-	Sigmoid
	MaxPool3D	1/1	$4 \times 1 \times 1$	1	-	-

Таблица 1: Архитектура нейронной сети UNET3d. Отказ от конкатенации активаций в пользу сложения обусловлен отсутствием потерь в качестве и ускорением работы нейронной сети.

	Слой	Число каналов (вход/выход)	Размер фильтра	Страйд	Доп. действия	Активация
Кодировщик	Conv2D_TimeDistributed	1/128	5×5	2	-	ReLU
	BatchNorm2D_TimeDistributed	128/128	-	-	-	-
	Conv2D_TimeDistributed	128/64	5×5	2	-	ReLU
	BatchNorm2D_TimeDistributed	64/64	-	-	-	-
LSTM	ConvLSTM	64/32	5×5	1	Dropout, 0.1 LayerNorm	ReLU
	ConvLSTM	32/64	5×5	1	Dropout, 0.1 LayerNorm	ReLU
	BatchNorm2D_TimeDistributed	64/64	-	-	-	-
Декодировщик	Conv2DTranspose_TimeDistributed	64/64	5×5	2	-	ReLU
	BatchNorm2D_TimeDistributed	64/64	-	-	-	-
	Conv2DTranspose_TimeDistributed	64/128	5×5	2	-	ReLU
	BatchNorm2D_TimeDistributed	128/128	-	-	-	-
	Conv2DTranspose_TimeDistributed	128/1	5×5	2	-	Sigmoid

Таблица 2: Архитектура нейронной сети ConvLSTMAE. Под использованием дропаута понимается применение этого механизма согласно статье Гала и Гарамани Gal and Ghahramani [2016]. LayerNorm применяется к изображению на каждом рекуррентном шаге ConvLSTM.

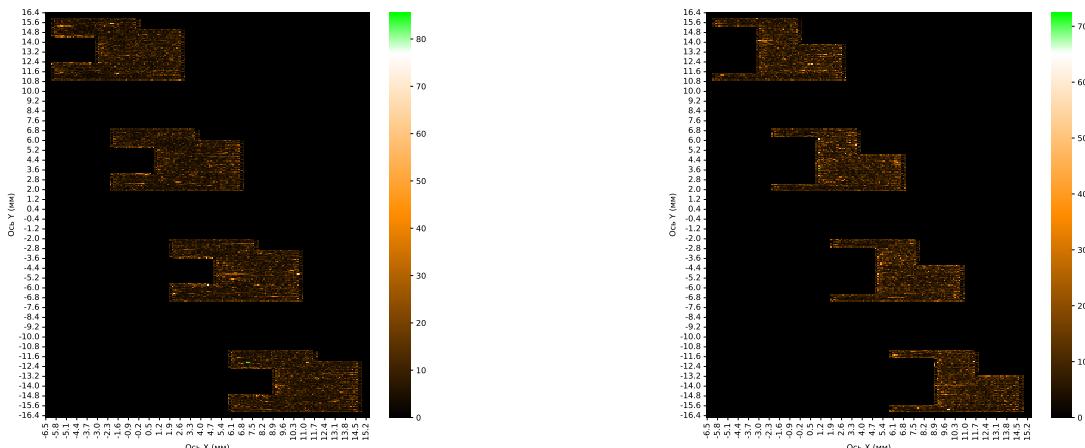
B Распределения функционала ошибки восстановления по координатам



(a) Среднее по координате значение функционала ошибки восстановления для тренировочного набора 210A.

(b) Среднее по координате значение функционала ошибки восстановления для валидационного набора 150A.

Рис. 11: Распределение функционала (1) по координате для модели ConvLSTMMAE, обученной на 210B с размером окна 2.



(a) Среднее по координате значение функционала ошибки восстановления для тренировочного набора 210A.

(b) Среднее по координате значение функционала ошибки восстановления для валидационного набора 150A.

Рис. 12: Распределение функционала (1) по координате для модели ConvLSTMMAE, обученной на 210B с размером окна 6.