

Univariate Time Series Missing Data Imputation using Pix2Pix GAN

Mauricio M. Almeida, João D. S. Almeida, Geraldo B. Junior, Aristófanés C. Silva, Anselmo C. Paiva

Abstract—The use of data is essential for the supply of business, scientific and other processes. Often the consumption of these data is hampered when there are sample losses. Aiming to recover values representative of these losses, there are several approaches for filling them. In this paper, we propose a new method for imputation of missing data that transforms time series into an image and thus performs imputation using the conditional generative adversarial network (cGAN) pix2pix GAN. The results of ASMAPE and MAE show that the network outperforms all methods in 50% of the datasets. It was also revealed that the proposed network can learn time series features and retain some advantages over traditional methods, such as imputing the data in its entirety and exploiting spatial and temporal features for imputation.

Index Terms—Time Series, Imputation, Pix2Pix network, cGAN.

I. INTRODUÇÃO

Uma série temporal (ST) é qualquer sequência de observações ordenadas no tempo, podendo ser discreta ou contínua. A principal característica dessa sequência é ela poder ser descrita por um processo estocástico, necessitando de modelagem estatística, como exemplo, *Random Walk* [1], processos autorregressivos [2], dentre outros.

Esses tipos de dados são extremamente importantes, pois por meio deles somos capazes de entender, prever e explicar processos caóticos de diversas áreas, a saber: economia [3], medicina [4], eventos naturais [5], entre outras. No entanto, quando se trata de dados de séries reais é muito provável que algumas amostras desses dados sejam perdidas durante sua aquisição, transferência ou armazenamento, dificultando seu estudo.

Quando uma série apresenta valores ausentes, esses valores faltosos podem ser preenchidos usando modelos de previsão. Além desses modelos de previsão, existem modelos específicos para imputação de dados. Alguns desses modelos propostos para abordagem de séries temporais são processos puramente matemáticos, tais como Interpolação Linear [6] e Média Móvel [7]. Outros modelos baseados em aprendizado de máquina também têm sido aplicado por Lakshminarayan et al. [8], além de trabalhos recentes como de Rani & Solanki [9], Hussain et al. [10] e Huamin et al. [11] que utilizam redes neurais para imputação de dados faltosos em diferentes tipos de séries temporais.

Assim, usando redes neurais, Guo et al. [12] propuseram um modelo de imputação por meio de redes generativas

adversárias (GANs) denominada MST-GAN (uma variação da GAN) e mostram que essa supera variações das GAN's como RGAN, métodos como interpolação cúbica e spline.

Por outro lado, os estudos expostos mostram as redes neurais apenas sendo utilizadas na imputação de dados multivariados, necessitando de estudos nesse mesmo sentido aplicados em séries univariadas. Assim, propõe-se a imputação em séries temporais por meio da conjugação dos métodos de reconstrução de imagem usando Pix2Pix GAN e transformação de séries 1D em imagens 2D. O presente estudo tem como contribuição uma nova abordagem para imputação de dados faltosos em séries temporais univariadas (STU) transformando as STU em imagens e recuperando os valores faltosos por meio da rede Pix2Pix GAN.

Este trabalho está organizado como segue: a Seção (II) descreve alguns estudos utilizando transformação de ST em imagem e usando redes neurais para imputação. A Seção III descreve os dados, as redes, as métricas de avaliação e o método proposto. A Seção V apresenta e discute os resultados, por fim, a Seção VI descreve as conclusões obtidas com o experimento bem como as perspectivas de pesquisas futuras.

II. TRABALHOS RELACIONADOS

Na exploração de características visuais, a área de visão computacional obteve sucesso ao ser utilizada em diversos contextos. Em Kawakura & Shibasaki [13] é proposto um sistema autônomo de entrega de materiais agrícolas. Por outro lado, Kortli et al. [14] fazem aplicações em veículos não tripulados, reconhecimento facial e de armas, controle de acesso a prédio e maquinários industriais. Essas aplicações extrapolam a área de processamento de imagens, pois exploram dados que originalmente representam pixels, mas na realidade representam relações espaciais do mundo real.

Wang & Oates [15] propuseram imputar valores em séries temporais a partir de redes neurais convolucionais. Para tanto, desenvolveram novos métodos de representar séries temporais em imagens, denotados por meio das técnicas *Gramian Angular Summation/Difference Field (GASF/GADF)* e *Markov Transition Field (MTF)*. Com isso, mostraram que o método proposto superou alguns dos melhores métodos existentes na área até então.

Por outro lado, em tarefa de classificação, Hatami et al. [16] utilizam a representação de séries 1D em imagens 2D por meio de gráficos de recorrência em espaço de fase. Com isso, mostraram que o método por eles proposto superou modelos que fazem uso de *Scale-invariant feature transform (SIFT)* e *Local binary patterns (LBP)* com classificador *Support-vector machine (SVM)*.

Contudo, Zhuang et al. [17] utilizando um método simples de representação de ST univariadas em imagem 2D, juntamente com rede neural convolucional, mostraram que o erro obtido nos experimentos são estáveis e inferiores ao de outros métodos testados na mesma base de dados de tráfego de veículos.

O método de Zhuang et al. [17] consiste em agrupar séries temporais em colunas, no qual o eixo horizontal denota os IDs de detectores ou seja, as séries de tráfego de veículos de uma estrada ou várias estradas que se cruzam. No eixo vertical tem-se os *timestamp*. Assim, com essa abordagem esse trabalho mostra que agrupar diferentes séries levando em conta uma lógica espaço temporal melhora os resultados das imputações em comparação com métodos tradicionais.

De outra forma, a proposta de Luo et al. [18] apresenta uma variação da rede *Gated Recurrent Unit (GRU)* denominada *Gated Recurrent Unit for data Imputation (GRUI)* que juntamente com bloco *Recurrent neural network (RNN)* compõe uma GAN. Com isso, ao realizarem imputações em séries multivariadas concluíram que a rede composta com GRUI geraram resultados melhores que GRU e *Long short-term memory (LSTM)*.

Abordando ST como imagem, Brophy et al. [19] remodelam a série temporal para imagens RGB rasterizadas, mostram que as redes GAN's oferecem excelente suporte a imputação de dados. Essas conclusões também são observadas em estudos recentes dos mesmos autores [20].

A abordagem desse estudo baseia-se inicialmente nos trabalhos de Oehmcke et al. [21], Pereira et al. [22] e Xia et al. [23]. Oehmcke et al. [21] demonstra a eficiência da rede Pix2Pix em criar imagens de satélites sem nuvens a partir de imagens existentes, Pereira et al. [22] mostra que essa rede é uma das mais eficientes na reconstrução de imagens, e Xia et al. [23] que usaram a mesma pra recuperar dados perdidos de imagens médicas.

Para usar a rede Pix2Pix foi necessário aprimorar um método para transformar séries temporais em imagens. Assim, o método proposto baseia-se, também, no trabalho de Zhuang et al. [17], pois se mostrou eficiente ao possibilitar a extração de características espaciais e temporais utilizando séries distintas.

Portanto, diferentemente dos trabalhos supracitados e buscando melhorar a tarefa de imputação de dados, a rede Pix2Pix foi adaptada para imputar dados faltosos em séries temporais univariadas, por meio da abordagem da transformação em imagem.

III. MATERIAIS E MÉTODO

A. Dataset

Neste estudo foram utilizados 9 datasets com dados de 7 áreas diferentes, e utilizou-se 50 amostras aleatórias do dataset LD201, totalizando 58 séries. Das subséries do dataset *Genetic and Evolutionary Computation Conference (GECCO)* utilizou-se apenas o campo Return Temperature. Os dados das séries utilizadas estão na faixa de valores de -8 a 235.625.

A Tabela I apresenta os datasets e suas principais informações. No dataset Return Temperature os valores originalmente faltosos foram substituídos pelo último valor válido.

Em todas as séries as imputações foram realizadas sobre as mesmas condições.

TABELA I
DESCRIÇÃO DOS DATASETS UTILIZADOS NO EXPERIMENTO

Dataset	Amostras por série	Origem
ElectricityLoadDiagrams 20112014 Data Set (LD2011)	140.256	[24]
Ammonia concentration (Ammo)	1.440	[25]
Distillate flowrate (Dest)	44.640	[25]
Electricity usage (Elet_usa)	2.712	[25]
Bike Sharing Dataset Data Set (Bike)	17.380	[26]
Parking Birmingham Data Set (Parkin)	35.718	[27]
daily_minimum_ temperatures_in_me (Daily)	3,650	[28]
Monthly Sunspots (Manch)	2.820	[29]
GECCO (Return)	606.838	[30]

B. Método Proposto

As séries temporais são ricas em informações que servem para auxiliar processos de decisão nas mais diversas áreas. No entanto, toda informação contida nas STs devem ser fieis aos valores reais coletados para que toda decisão tomada baseada nessas informações sejam mais acertada possível, mas nem sempre tem-se todas as informações intactas. Essa pesquisa levanta a hipóteses de que a exploração da relação espacial e temporal das STU podem ser mais eficientes do que considerar apenas a ordem temporal dos dados. Além disso, levanta a hipótese que as redes GANS são eficientes na exploração dessas características.

O método dessa pesquisa pode ser descrito em 3 etapas: A) aquisição de séries univariadas e pré-tratamento, B) Transformar séries temporais em Imagens 2D, C) Treinamento da GAN para imputar valores faltosos.

1) *Aquisição e pré-tratamento*: Há diversas bases de dados públicas de séries temporais, a maioria dessas são multivariadas, restando uma pequena parte univariada. Dessa pequena parte univariada, poucas são de dados reais e menos ainda com número de amostras superior há 2000 elementos. A quantidade de amostras é essencial para este trabalho, visto que uma imagem 100x100 pixels já possui 10.000 elementos. Por tanto, para que se tenha uma quantidade mínima de amostras que possam oferecer características espaciais, escolheu-se as séries que possuam pelo menos 1024 elementos, refletindo assim em uma imagem 32x32.

Além disso, realizou-se a padronização da disposição das séries na orientação vertical, bem como exclui-se as séries que não possuam marcação temporal.

2) *Método de transformação de séries temporais em imagem 2D*: Para se construir uma imagem que forneça características espaço-temporal a partir de uma série temporal, nos baseamos em Zhuang et al. [17].

As imagens são criadas a partir do empilhamento de trechos das séries que contém data/período temporal semelhantes. Suponhamos dados de vendas bimestrais de veículos em uma determinada empresa. Aplicando-se o método proposto, dispõe-se os trechos sobre um mesmo padrão onde cada linha e coluna representam, respectivamente, um ano e um bimestre. Pode-se visualizar este exemplo na Fig. 1, na qual reorganizamos a série que contém dados de 2000 a 2020, formando assim uma imagem 2D de 6x6 pixels.

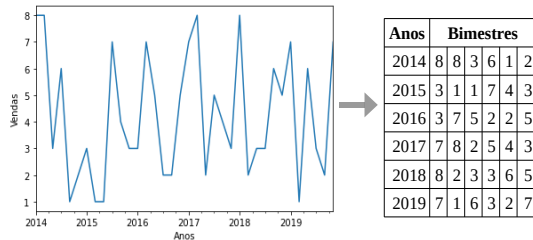


Fig. 1. Exemplo de empilhamento de janelas.

3) *Pix2Pix GAN*: A rede Pix2Pix [31] foi desenvolvida com base no cGAN [32]. Dentre outras aplicações, essa rede é usada na reconstrução de objetos a partir de bordas, fotos de mapas de síntese e colorização de imagens. Essa rede possui dois principais componentes o gerador e o decodificador. O gerador tem objetivo de gerar uma imagem a partir de um esboço que no caso dessa pesquisa é uma imagem com valor faltoso. De outra forma, o decodificador tem o objetivo de aprender a diferenciar uma imagem real de uma imagem gerada pelo gerador.

Realizamos mudanças no gerador original da rede. A arquitetura proposta para o gerador da Pix2Pix pode ser visualizada na Fig. 2a. Essa mudança foi necessária devido a entrada original da rede ser 256x256 e a nova entrada ser 32x32, resultando assim em uma versão mais compacta da rede Pix2Pix [31]. Além disso, foi alterada a função de ativação do último bloco de Convolução de tangente hiperbólica para sigmoide. Essa mudança trouxe melhores resultados, pois como as imagens de entrada da rede normalizadas pertencem ao conjunto [0,1] e, esse conjunto por sua vez é o contradomínio da função sigmoide somado a isso a função de custo da rede é uma composição do Erro Médio Absoluto, uma função positiva. Portanto, a sigmoide é mais adequada.

Por outro lado, o discriminador foi adaptado alterando-se o tamanho dos *kernels* de 4 para 2 e *strides* para 2, pois dada uma imagem de resolução pequena, tais alterações possibilitaram maior extração de características das séries temporais. A estrutura do gerador pode ser visualizada na Fig. 2a e a estrutura da rede em alto nível, pode ser visualizada na Fig. 3.

A rede proposta possui 6.569.986 parâmetros treináveis e os métodos comparados possuem complexidade $O(n)$ a $O(n^2)$. O tempo necessário para o treinamento da rede com 500 mil passos (equivalente a 166 épocas) leva um tempo inferior a 2 horas. A rede base utilizada para a implementação está disponível no *TensorFlow* [33]. O código-fonte da versão proposta, bem como a base de dados está disponível no *GitHub* [34].

IV. EXPERIMENTO

Para realização dos experimentos a divisão dos dados na transformação T1 foi de 83%, 9% e 8% para treino, validação e teste, respectivamente, e na mesma ordem 64%, 18% e 18% para transformação T2. Não foi possível utilizar todos os datasets para o teste, pois o número de observações da ST de alguns destes não ultrapassaram 2.048 ou não possuem marcação temporal para se utilizar T2. As imagens de entrada da rede são normalizadas utilizando normalização *Max-Min*.

Assim, foram realizadas imputações nas amostras de teste das duas abordagens utilizando métodos tradicionais disponíveis na biblioteca *pandas* e uma implementação do K-ésimo Vizinho mais Próximo (k-nn) baseado em Flores et al. [35]. Os métodos utilizados foram Média (M), Mediana (MD), Interpolações (linear (IL), cúbica (IC), slinear (IS), akima (IA), polinomial quádrupla (IP4), polinomial spline grau 5 (IP5)) e média móvel com janelamento de 3 elementos (MM3).

Contudo, alguns desses métodos nem sempre conseguem imputar os dados faltosos em sua totalidade, a exemplo temos as interpolações spline, akima, slinear, linear e polinomiais. Essas, não conseguem calcular a derivada para dois valores faltosos adjacentes. Diferente das interpolações, o 4-NN não imputa valores em lacunas com mais de 4 elementos faltosos, resultando, assim, na falta de previsão desses valores. Por isso, os dados não imputados foram desconsiderados na avaliação dos resultados e assim se aferiu as métricas nas mesmas instâncias para as quais todos os métodos conseguiram imputar.

Em seguida, substituiu-se aleatoriamente valores de cada imagem, simulando valores faltosos a uma taxa de 40 % em cada série. Visando fornecer o máximo de informação para a rede, testou-se a substituição por um valor que corresponde ao [mínimo-1], [máximo+1], zero e média. Os melhores resultados foram obtidos alterando o valor original pela média de toda a imagem, pois esses valores possuem uma relação mínima com os demais valores, sendo também ponto de extração de características, ou seja, carregam informação estatística de tendência central dos dados e não somente preenchimento de espaço, como seria o caso da substituição por zero. Assim, dada uma imagem contendo 32 janelas de 32 amostras, cada uma das 32 janelas (subséries) teve 13 valores substituídos pela média.

V. RESULTADOS E DISCUSSÃO

Conforme descrito na Seção III-B, uma imagem alterada é inserida na rede com objetivo de recuperar os valores originais (pixel) a partir do esboço recebido. A principal vantagem de se trabalhar a imputação com a abordagem por imagens, está no fato das redes neurais convolucionais usarem atributos da série em tempos passados e futuros aos valores faltosos, bem como atributos de séries vizinhas. Sendo assim, são mais sensíveis ao tempo, sazonalidade e tendência.

Nos exemplos da Fig. 4, nota-se que a diferença entre os valores originais e imputados é muito sutil, indicando que a rede Pix2Pix é capaz de aprender e gerar características de séries temporais. Os experimentos realizados nesse estudo

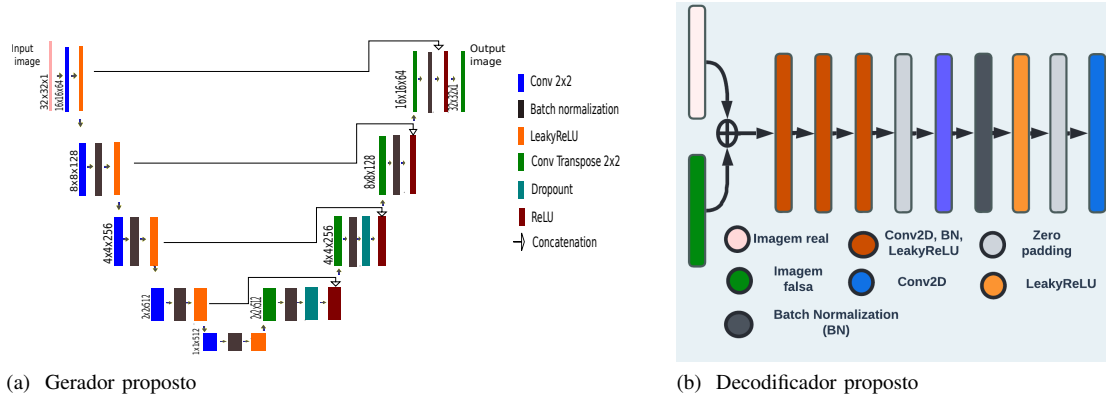


Fig. 2. Componentes propostos para rede Pix2Pix.

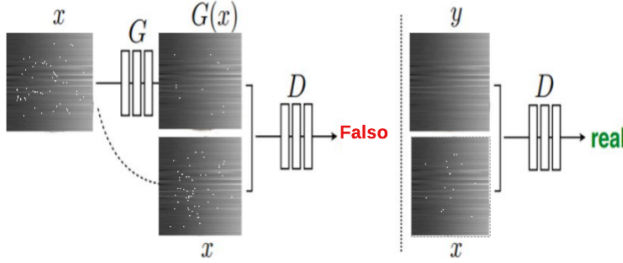


Fig. 3. Estrutura da rede Pix2Pix em alto nível. Adaptada de [31].

X é imagem com valores faltosos, y imagem real, G(X) gerador da rede e D o discriminador da rede.

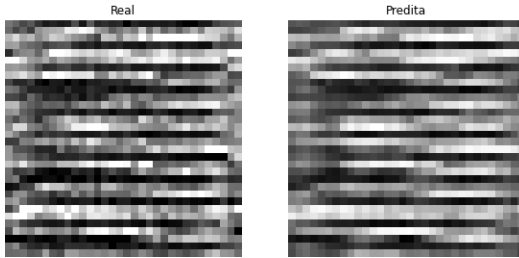


Fig. 4. Imagem real (REAL) e imagem gera pela rede (Predita).

foram avaliados em função do Erro Médio Absoluto (MAE) (1) e do Erro Médio Quadrático (RMSE) (2), métricas já consagradas no estudo de STU.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

Onde y_i é o valor real e \hat{y}_i o valor predito.

Para uma avaliação percentual, propôs-se a métrica Adapted Symmetric Mean Absolute Percentage Error (ASMAPE) que consiste em uma adaptação do SMAPE [36] convencional para resolver os casos de indeterminação. O SMAPE e ASMAPE

(3) são equivalentes, pois a primeira desconsidera as indeterminações, enquanto a segunda apenas as altera para zero. Decidiu-se evidenciar tal abordagem, pois na literatura atuais abordagens não são explicitadas ao leitor.

$$ASMAPE = \frac{1}{n} \sum_{i=1}^n \begin{cases} 0, & \text{se } y_i = \hat{y}_i = 0 \\ \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}, & \text{caso contrário} \end{cases} \quad (3)$$

onde y_i é o valor real e \hat{y}_i o valor predito.

As Tabelas II, III e IV mostram o MAE, RMSE e ASMAPE, respectivamente, em cada dataset de teste bem como por método. Nessas são mostrados o resultados obtidos pela rede que foi treinada inicialmente com a transformação T1 e os pesos usados como base para o treinamento com T2.

Nota-se que a rede Pix2Pix supera todos os métodos nos datasets Bike, Dayle e Manch. Nos outros datasets, apresenta resultado equivalente a outros métodos tradicionais como interpolação cúbica, spline e média móvel, e nestes, supera o 4-nn, média e mediana.

Embora a rede não supere todos os métodos, ela apresenta a vantagem de imputar todos os valores faltosos. A quantidade de valores não imputados por cada métodos nas transformações T1 e T2 são semelhantes com uma leve desvantagem da T1 sobre e T2. Apenas a Pix2Pix, a média e a mediana imputaram todos os valores, os demais métodos deixam de imputar valores na ordem de 10^3 a 10^5 . Para se evidenciar essas vantagens foram realizados dois experimentos principais. No primeiro (Transformação T1), as séries foram subdivididas em subséries contendo 1024 elementos, cada. Em seguida, cada subsérie é transformada em uma matriz 2D, através de um processo de empilhamento, resultando em imagens 2D com dimensão 32 x 32 pixels. O segundo experimento (Transformação T2) é realizado por meio da transformação proposta diferencia-se do primeiro no que diz respeito a organização, pois as subséries são empilhadas considerando o alinhamento temporal dos dados. A Fig. 1 (Seção III-B) ilustra o processo de representação da série em imagem por meio da Transformação T2.

Quanto ao RMSE (Tabela III), nota-se que em alguns datasets, é bastante alto. Esses valores são justificados pela presença de amostras difíceis de aprender, isso fica evidente

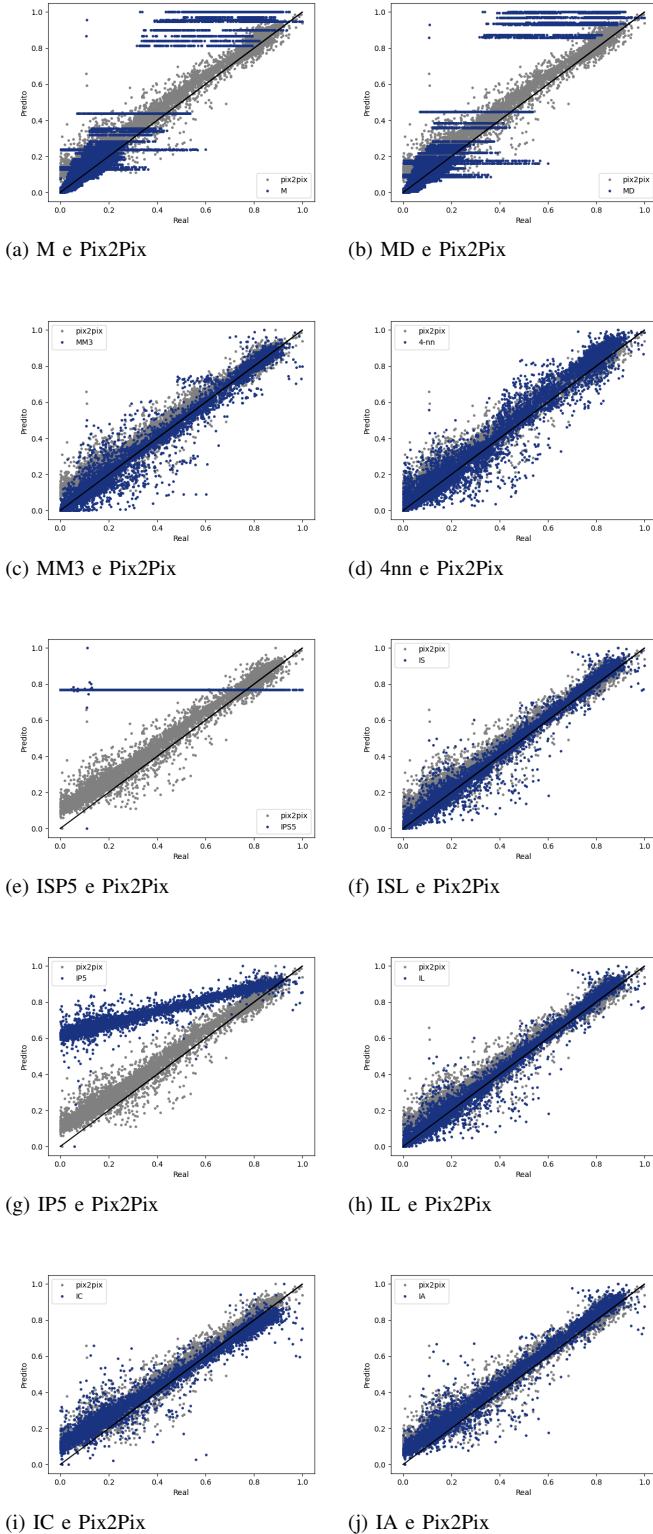


Fig. 5. Comparação normalizada entre previsão de cada método com a Pix2Pix.

na Fig. 5g, que tem para os valores reais de 0.1 a 0.6 diversas amostras com valores deslocados da diagonal. Pode contribuir para isso, a alta taxa de valores faltosos inseridos,

bem como a escala dos dados, pois essa métrica é sensível à escala. Estes casos podem ser visualizados na Fig. 5. Ainda na Fig. 5 e Tabela III, nota-se que tais dificuldades também são enfrentadas pelos métodos utilizados, visto que há grande dispersão de valores para os mesmos.

Os resultados nos dados de teste mostram que o método de transformação de séries temporais, considerando a ordem temporal, pode reduzir o erro, pois nas Tabelas II, III e IV observa-se que tanto a rede quanto os métodos tradicionais se saíram melhores em alguns dos datasets com o método proposto, T2.

A menor taxa de amostras nos experimentos que considera a ordem espaço temporal pode ter ocasionado um desempenho ligeiramente inferior em T2 para a rede e os métodos. Essa redução de amostras já era prevista conforme descrito na Seção III.

VI. CONCLUSÃO

Este artigo propôs uma nova abordagem para imputação de dados faltosos em séries temporais utilizando rede *Pix2Pix* e uma nova maneira de transformar ST's em imagens. Mostramos que a rede *Pix2Pix* é capaz de aprender e gerar ótimos resultados que superaram todos os métodos em pelo menos 50% dos datasets utilizados, além de se equiparar aos melhores métodos no LD2011. Somado a isso, evidencia-se a vantagem sobre todos os métodos não superados de imputar todos os valores faltosos sem exceção.

Os resultados alcançados nos experimentos evidenciam que a rede aprende melhor o padrão dos dados quando a organização da imagem segue uma lógica de organização espaço-temporal, o que mostra que a nossa proposta de transformação de séries em imagem explora as principais características de uma série que são sazonalidade, tendência e ciclo. Além disso, conclui-se que o uso de redes neurais convolucionais apresentam a vantagem de explorar características e relacionar observações em tempos diferentes sem maiores custos.

Por fim, sabe-se que o assunto ainda não se exauriu. Assim, para pesquisas futuras sugere-se utilizar o mesmo pipeline com outras redes como LSTM, RNN, 1D-CNN, Support Vector Regression (SVR) e com métodos diferentes como Facebook Prophet 37 e missForest [38], pois mostramos que T2 pode contribuir de forma significativa para melhores resultados.

Pode-se ainda analisar o desempenho da rede *Pix2Pix* por tipo de série. Para isso, pode-se realizar os mesmos experimentos com grupos de séries do mesmo nicho ou separadas por grupos homogêneos nas características de sazonalidade, ciclo e tendência. Assim, esses resultados também podem contribuir para a linha de pesquisa de classificação de séries temporais.

VII. ACKNOWLEDGMENTS

Os autores agradecem o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão (FAPEMA).

TABELA II
MAE POR DATASET COM A TRANSFORMAÇÃO T1 E T2

Datasets	M	MD	IL	IC	IS	IA	IP5	ISP5	4-NN	MM3	PIX2PIX	
T1	Parkin	141.50	96.00	16.07	15.84	16.07	13.03	22.17	22.00	42.53	25.00	37.28
	LD2011	19.63	15.76	3.62	4.28	3.62	3.70	5.20	4.66	5.55	4.47	4.27
	Return	10.31	9.30	0.10	0.11	0.10	0.10	0.13	0.49	0.80	0.35	2.26
	Bike	78.91	69.00	27.54	27.58	27.54	23.74	36.12	36.93	53.83	41.00	24.49
	Daily	2.65	2.60	1.50	1.57	1.50	1.54	2.16	1.79	1.50	1.80	1.41
	Manch	2.45	2.30	1.64	1.91	1.64	1.76	2.30	1.98	1.75	2.03	1.55
	Média	42.58	32.49	8.41	8.55	8.41	7.31	11.35	11.31	17.66	12.44	11.88
T2	Parkin	174.18	148.50	22.33	23.85	22.33	17.87	26.97	26.94	65.21	42.00	63.85
	LD2011	17.33	14.52	3.44	4.04	3.44	3.51	4.87	4.37	5.15	4.26	3.87
	Return	9.83	8.75	0.13	0.14	0.13	0.12	0.18	0.54	0.71	0.30	2.10
	Bike	130.21	115.00	49.82	51.90	49.82	45.73	63.27	63.66	90.29	61.00	44.41
	Daily	3.03	3.00	1.56	1.94	1.56	1.64	2.23	1.64	1.44	1.80	1.30
	Manch	3.17	3.10	1.71	2.02	1.71	1.76	2.60	1.83	1.70	2.00	1.29
	Média	56.29	48.81	13.17	13.98	13.17	11.77	16.69	16.50	27.42	18.56	19.47

TABELA III
RMSE POR DATASET COM A TRANSFORMAÇÃO T1 E T2

Datasets	M	MD	IL	IC	IS	IA	IP5	ISP5	4-NN	MM3	PIX2PIX	
T1	Parkin	224.59	235.13	88.86	91.42	88.86	85.73	125.91	22916.28	137.73	110.47	107.23
	LD2011	54.68	59.25	13.15	15.95	13.15	13.52	31.91	>10e5	20.91	17.03	16.64
	Return	14.71	14.95	2.73	3.16	2.73	2.75	4.14	40.42	4.41	3.97	6.22
	Bike	123.83	125.30	81.88	95.41	81.88	83.90	154.82	1136.06	101.57	103.89	69.63
	Daily	3.63	3.64	2.77	3.46	2.77	2.92	5.63	3.60	2.69	3.21	2.70
	Manch	3.42	3.42	2.82	3.66	2.82	2.96	12.26	3.75	2.74	3.41	2.55
	Média	70.81	73.62	32.04	35.51	32.04	31.96	55.78	>10e5	45.01	40.33	34.16
T2	Parkin	206.44	212.21	72.06	77.67	72.06	70.13	92.85	91.19	111.71	97.37	95.85
	LD2011	52.20	56.87	13.09	16.18	13.09	13.59	27.74	>10e5	20.94	17.35	15.63
	Return	13.56	13.58	2.73	2.92	2.73	2.71	3.39	3.39	4.45	3.88	5.38
	Bike	190.30	193.67	117.88	134.97	117.88	122.74	177.88	178.76	157.65	151.42	117.15
	Daily	4.03	4.01	2.38	3.18	2.38	2.59	4.82	3.38	2.35	2.95	2.33
	Manch	4.09	4.09	2.64	3.31	2.64	2.73	4.83	2.99	2.78	3.11	2.53
	Média	78.44	80.74	35.13	39.71	35.13	35.75	51.92	>10e5	49.98	46.01	39.81

TABELA IV
ASMAPE POR DATASET COM A TRANSFORMAÇÃO T1 E T2

Datasets	M	MD	IL	IC	IS	IA	IP5	ISP5	4-NN	MM3	PIX2PIX	
T1	Parkin	23.26	21.61	6.88	8.41	6.88	6.53	10.61	10.96	12.29	9.77	9.46
	LD2011	14.25	13.36	4.86	6.39	5.31	5.50	7.73	6.29	6.30	5.95	5.76
	Return	13.94	13.55	0.80	0.90	0.80	0.70	1.21	1.49	2.17	1.30	4.37
	Bike	39.49	40.20	23.57	27.54	23.57	23.65	33.33	33.82	34.58	31.58	22.46
	Daily	13.19	13.18	9.36	11.23	9.36	9.84	15.70	11.68	9.40	11.33	9.16
	Manch	12.36	12.33	9.49	11.31	9.49	9.90	15.66	12.64	9.84	11.71	8.99
	Média	19.42	19.04	9.16	10.96	9.24	9.35	14.04	12.81	12.43	11.94	10.03
T2	Parkin	18.18	17.50	4.98	6.01	4.98	4.82	7.46	7.31	9.80	8.09	8.09
	LD2011	14.17	13.63	3.59	4.46	3.69	3.81	5.80	5.35	5.31	4.56	4.20
	Return	10.85	10.45	0.89	1.03	0.89	0.80	1.32	1.58	2.26	1.34	3.73
	Bike	37.99	38.28	23.73	28.75	23.73	24.06	33.04	32.90	33.04	31.33	22.75
	Daily	15.60	15.58	9.21	12.22	9.21	9.99	15.67	11.39	9.62	11.30	9.02
	Manch	15.69	15.62	10.41	13.67	10.41	10.82	18.68	11.97	10.47	12.58	9.24
	Média	18.75	18.51	8.80	11.02	8.82	9.05	13.66	11.75	11.75	11.53	9.51

REFERENCES

834, 2008.

- [1] E. A. Codling, M. J. Plank, and S. Benhamou, "Random walk models in biology," *Journal of the Royal society interface*, vol. 5, no. 25, pp. 813–834, 2008.
- [2] R. S. Tsay, "Testing and modeling threshold autoregressive processes," *Journal of the American statistical association*, vol. 84, no. 405, pp. 231–

- 240, 1989.
- [3] Y. Zhao and Y. Zhou, "Measurement method and application of a deep learning digital economy scale based on a big data cloud platform," *Journal of Organizational and End User Computing (JOEUC)*, vol. 34, no. 3, pp. 1–17, 2022.
- [4] D. Jarrett, J. Yoon, I. Bica, Z. Qian, A. Ercole, and M. v. d. Schaar, "Clairvoyance: A pipeline toolkit for medical time series.," 2021.
- [5] G. Papacharalampous, H. Tyrallis, and D. Koutsoyiannis, "Predictability of monthly temperature and precipitation using automatic time series forecasting methods," *Acta Geophysica*, vol. 66, no. 4, pp. 807–831, 2018.
- [6] H. Aboussaid, "The effect of ethnic fractionalization on economic development. a multilevel analysis," 2021.
- [7] W. A. Fuller, *Introduction to statistical time series*. John Wiley & Sons, 2009.
- [8] K. Lakshminarayan, S. A. Harp, R. P. Goldman, T. Samad, *et al.*, "Imputation of missing data using machine learning techniques.," in *KDD*, vol. 96, 1996.
- [9] S. Rani and A. Solanki, "Data imputation in wireless sensor network using deep learning techniques," in *Data analytics and management*, pp. 579–594, Springer, 2021.
- [10] S. N. Hussain, A. A. Aziz, M. Hossen, N. A. A. Aziz, G. R. Murthy, F. B. Mustakim, *et al.*, "A novel framework based on cnn-lstm neural network for prediction of missing values in electricity consumption time-series datasets.," *Journal of Information Processing Systems*, vol. 18, no. 1, 2022.
- [11] T. Huamin, D. Qiuqun, and X. Shanzhu, "Reconstruction of time series with missing value using 2d representation-based denoising autoencoder," *Journal of Systems Engineering and Electronics*, vol. 31, no. 6, pp. 1087–1096, 2020.
- [12] Z. Guo, Y. Wan, and H. Ye, "A data imputation method for multivariate time series based on generative adversarial network," *Neurocomputing*, vol. 360, pp. 185–197, 2019.
- [13] S. Kawakura and R. Shibasaki, "Deep learning-based self-driving car: Jetbot with nvidia ai board to deliver items at agricultural workplace with object-finding and avoidance functions," *European Journal of Agriculture and Food Sciences*, vol. 2, no. 3, 2020.
- [14] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, p. 342, 2020.
- [15] Z. Wang and T. Oates, "Imaging time-series to improve classification and imputation," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.
- [16] N. Hatami, Y. Gavet, and J. Debayle, "Classification of time-series images using deep convolutional neural networks," in *Tenth international conference on machine vision (ICMV 2017)*, vol. 10696, p. 106960Y, International Society for Optics and Photonics, 2018.
- [17] Y. Zhuang, R. Ke, and Y. Wang, "An innovative method for traffic data imputation based on convolutional neural network," *IET Intelligent Transport Systems*, vol. 13, 11 2018.
- [18] Y. Luo, X. Cai, Y. Zhang, J. Xu, *et al.*, "Multivariate time series imputation with generative adversarial networks," *Advances in neural information processing systems*, vol. 31, 2018.
- [19] E. Brophy, Z. Wang, and T. E. Ward, "Quick and easy time series generation with established image-based gans," *arXiv preprint arXiv:1902.05624*, 2019.
- [20] E. Brophy, Z. Wang, Q. She, and T. Ward, "Generative adversarial networks in time series: A survey and taxonomy," *arXiv preprint arXiv:2107.11098*, 2021.
- [21] S. Oehmcke, T.-H. K. Chen, A. V. Prishchepov, and F. Gieseke, "Creating cloud-free satellite imagery from image time series with deep learning," in *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data*, pp. 1–10, 2020.
- [22] R. C. Pereira, M. S. Santos, P. P. Rodrigues, and P. H. Abreu, "Reviewing autoencoders for missing data imputation: Technical trends, applications and outcomes," *Journal of Artificial Intelligence Research*, vol. 69, pp. 1255–1285, 2020.
- [23] Y. Xia, L. Zhang, N. Ravikumar, R. Attar, S. K. Piechnik, S. Neubauer, S. E. Petersen, and A. F. Frangi, "Recovering from missing data in population imaging—cardiac mr image imputation via conditional generative adversarial nets," *Medical Image Analysis*, vol. 67, p. 101812, 2021.
- [24] A. T. Elergone, "ElectricityLoadDiagrams20112014 Data Set," 2020.
- [25] K. Dunn, "OpenMV.net Datasets." <https://openmv.net/>, 2018.
- [26] H. Fanaee-T and J. Gama, "Event labeling combining ensemble detectors and background knowledge," *Progress in Artificial Intelligence*, pp. 1–15, 2013.
- [27] D. H. Stolfi, E. Alba, and X. Yao, "Predicting car park occupancy rates in smart cities," in *International Conference on Smart Cities*, pp. 107–117, Springer, 2017.
- [28] KANKANA, "Daily minimum temperatures in me."
- [29] J. Brownlee, "Machine learning datasets." [urlhttps://raw.githubusercontent.com/jbrownlee/Datasets/master/monthly-sunspots.csv](https://raw.githubusercontent.com/jbrownlee/Datasets/master/monthly-sunspots.csv), dez 2020. accessed: 20-12-2021.
- [30] S. Moritz, M. Friese, A. Fischbach, C. Schlitt, and T. Bartz-Beielstein, "GECCO Industrial Challenge 2015 Dataset: A heating system dataset for the 'Recovering missing information in heating system operating data' competition at the Genetic and Evolutionary Computation Conference 2015, Madrid, Spain," May 2015.
- [31] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.
- [32] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.
- [33] "<https://github.com/mauriciomoraisa/univariate-time-series-missing-data-imputation-using-pix2pix-gan.git>,"
- [34] "<https://www.tensorflow.org/tutorials/generative/pix2pix>,"
- [35] A. Flores, H. Tito, and C. Silva, "Local average of nearest neighbors: Univariate time series imputation," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 8, pp. 45–50, 2019.
- [36] V. Kreinovich, H. T. Nguyen, and R. Ouncharoen, "How to estimate forecasting quality: A system-motivated derivation of symmetric mean absolute percentage error (smape) and other similar characteristics," 2014.
- [37] M. Daraghme, A. Agarwal, R. Manzano, and M. Zaman, "Time series forecasting using facebook prophet for cloud resource management," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, IEEE, 2021.
- [38] S. Zhang, L. Gong, Q. Zeng, W. Li, F. Xiao, and J. Lei, "Imputation of gps coordinate time series using missforest," *Remote Sensing*, vol. 13, no. 12, p. 2312, 2021.



Mauricio M. Almeida Received a degree in Mathematic from the Federal Institute of Education, Science and Technology of Maranhão (IFMA) (2020), He is currently a Master's student in Computer Science at the Federal University of Maranhão conducting research under the guidance of Prof. Dr. João D. S. de Almeida on the issue of imputation of faulty data in time series.



João D. S. de Almeida Received a degree in Computer Science from the Federal University of Maranhão (UFMA) (2007), a master's degree in Electrical Engineering from UFMA (2010), and a Ph.D. in Electrical Engineering from UFMA (2013). He is currently an Associate Professor I at UFMA. He coordinates the Vision and Image Processing Laboratory (VipLab-UFMA). He has experience in Computer Science, working mainly on the following topics: image processing, machine learning, ophthalmic medical images, and time series.



Geraldo B. Junior Received an undergraduate degree in Computer Science, a Master's degree in Electrical Engineering with emphasis on Computer Science, and PhD in Electrical Engineering with emphasis on Computer Science, all held at the Federal University of Maranhão (UFMA). He is an Associate Professor I at UFMA, a permanent member of the Post-graduation Programs of Master in Computer Science (PPGCC/UFMA) and Ph.D. in Computer Science / Association UFMA-UFPI. Has experience in Computer Science, working mainly on

the following topics: computer vision, machine learning, deep learning, and medical image processing.



Aristófanés C. Silva Received a bachelor's degree in Computer Science, a master's degree in Electrical Engineering from the Federal University of Maranhão (UFMA), and a PhD in Computer Science from the Pontifical Catholic University of Rio de Janeiro. He is currently a Full Professor at UFMA. He has experience in Computer Science, with emphasis on Graphic Processing (Graphics), working mainly on the following topics: medical imaging and artificial intelligence.



Anselmo C. Paiva Received a BSc in civil engineering from Maranhão State Univeristy -Brazil in 1990; an MSc in civil engineering-Structures in 1993; and a PhD in Informatics from the Pontifical Catholic University of Rio de Janeiro – Brazil in 2002. He is currently a Full Professor at the Informatics Department at the Federal University of Maranhão -Brazil. His current interests include medical image processing, geographical information systems and scientific visualization. He is the coordinator of the NCA-UFMA Applied Computing Center. Has experience in Computer Science, with emphasis on Graphics Processing, working mainly on the following topics: Virtual and Augmented Reality, Computer Graphics, GIS, Medical Image Processing and Volumetric Visualization. He is a member of SBC (Brazilian Computer Society) and ACM (Association for Computing Machinery)