

INSTITUTO MILITAR DE ENGENHARIA - IME
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

DescNet — Rede Neural Convolucional de Descritores: uma abordagem para o fechamento de *loop* em *Visual SLAM*



Alexandra Miguel Raibolt da Silva

Paulo Fernando Ferreira Rosa
(Orientador)

Alberto Torres Angonese
(Co-orientador)



Sumário

- Introdução
- Conceitos Básicos e Estado da Arte
- Trabalhos Relacionados
- Problema Proposto
- Experimentos e Resultados
- Considerações Finais
- Referências

Resumo

- Rede Convolucional Recorrente de Longo Prazo;
- Rede Neural Convolucional de Descritores (DescNet);
- Localização e Mapeamento Simultâneos baseado em visão;
- Detecção de fechamento de *loop*.

O problema de SLAM

Pode ser caracterizado pela resolução de tarefas complexas, tais quais:

- Mapeamento;
- Localização.

Estas duas tarefas não podem ser realizadas de forma independente, isto é, elas atuam de forma complementar.

Influência da incerteza na localização e mapeamento



Figura: Influência da incerteza na localização e mapeamento.

Fonte: [1] apud [2].

Influência da incerteza na localização e mapeamento

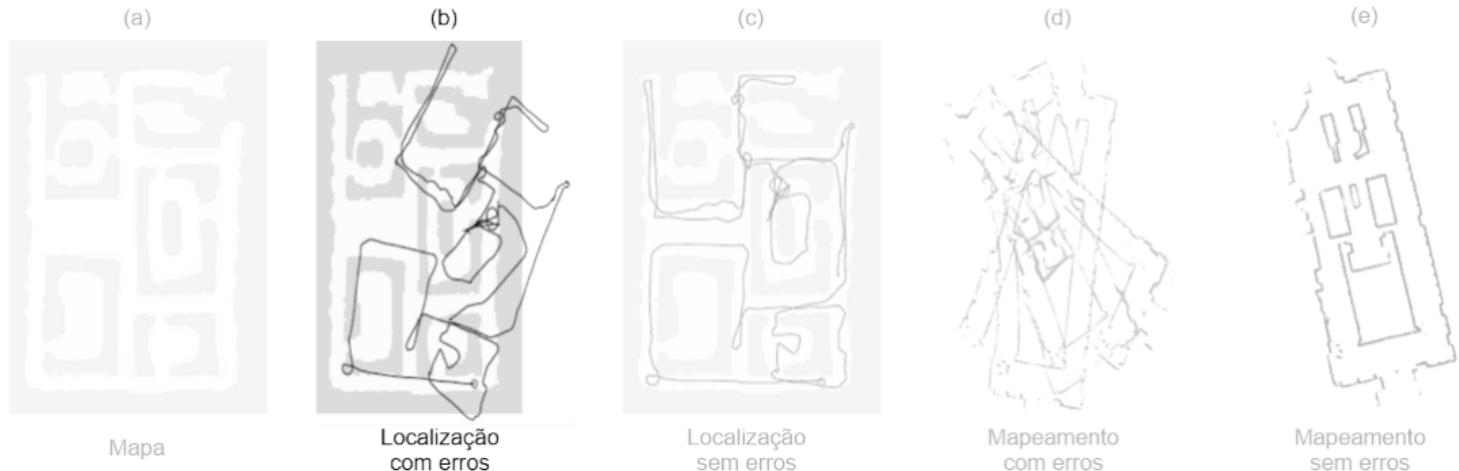


Figura: Influência da incerteza na localização e mapeamento.

Fonte: [1] apud [2].

Influência da incerteza na localização e mapeamento

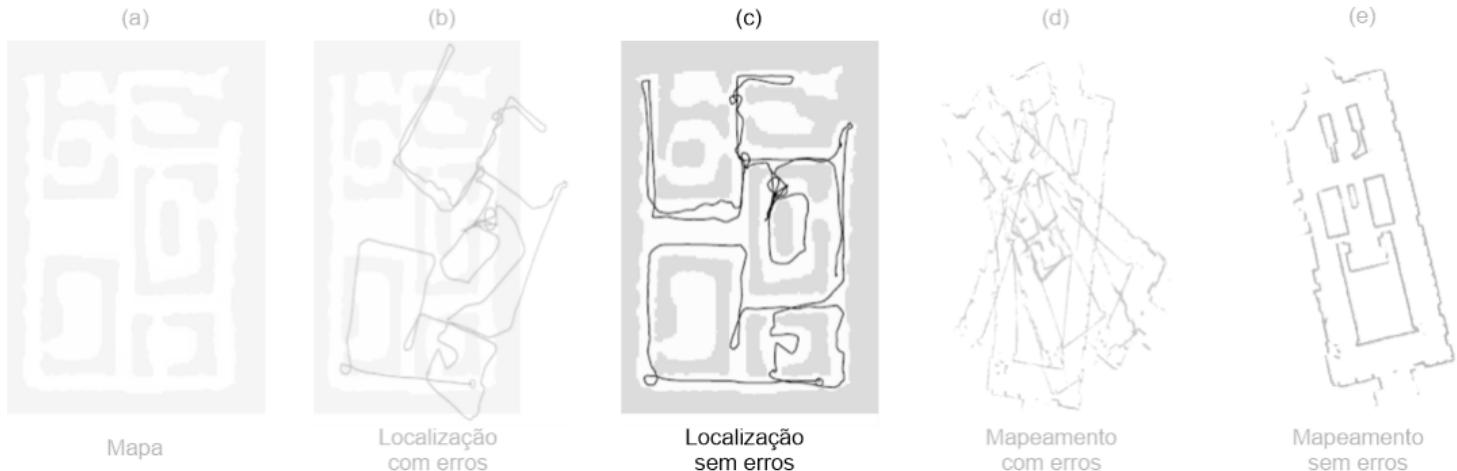


Figura: Influência da incerteza na localização e mapeamento.

Fonte: [1] apud [2].

Influência da incerteza na localização e mapeamento



Figura: Influência da incerteza na localização e mapeamento.

Fonte: [1] apud [2].

Influência da incerteza na localização e mapeamento

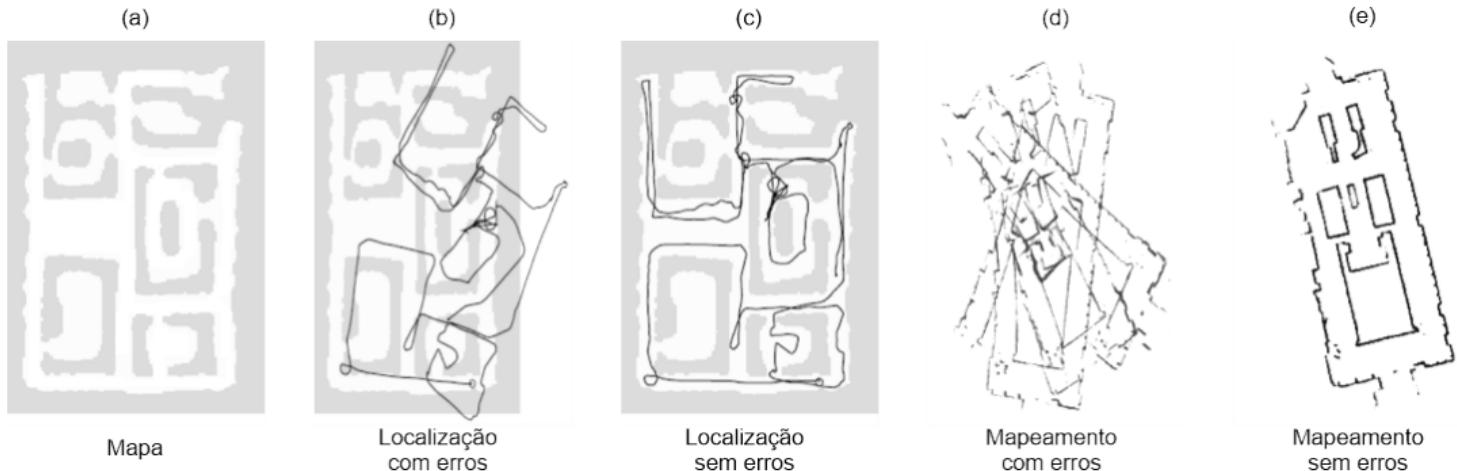


Figura: Influência da incerteza na localização e mapeamento.

Fonte: Adaptado de [1] apud [2].

Odometria

Métodos de Odometria:

- Roda;
- Sistema de posicionamento global (GPS);
- Sistema de navegação inercial (INS);
- Câmeras óticas.

Odometria Visual

<https://www.youtube.com/watch?v=trbBh8Rjc4s>

Odometria Visual + SLAM = VSLAM

Vantagens VSLAM:

- Utilização de câmeras;
- Alternativa barata a sensores e lasers.

Técnicas para solução de VSLAM:

- Redes Neurais Convolucionais;
- Redes Neurais Recorrentes.

Detecção de fechamento de loop

<https://www.youtube.com/watch?v=OV6wNr62nqQ>

Detecção de fechamento de loop

Técnica comumente explorada:

- Saco de Palavras Visuais.

Técnica pouco explorada:

- Aprendizado Profundo.

Motivação

Buscamos:

- Alternativas baratas;
- Robustas;
- Desempenho computacional eficiente.

Justificativa

IFR e AI Index

- Defesa;
- Operações militares;
- Busca;
- Resgate e salvamento;
- Logística;
- Serviços domésticos;
- Limpeza e desinfecção.

COVID-19

- Crise pandêmica global.



Figura: Rôbo de logística.



Figura: Rôbos de desinfecção.

Justificativa

Atender as demandas dos diversos setores de robôs de serviços, tanto no âmbito civil quanto militar.

Demandas estas, evidenciadas cada vez mais em meio a sociedade, e pelas iniciativas e estratégias estabelecidas por políticas (ao nível governamental e intergovernamental) direcionadas à Inteligência Artificial.

Objetivo

Integração de um modelo híbrido de arquitetura de Rede Neural Artificial — a Rede Convolucional Recorrente de Longo Prazo com blocos DescNet com um sistema de Localização e Mapeamento Simultâneos baseado em visão capaz de solucionar o problema de Detecção de fechamento de *loop*.

Detecção e Descrição de Características

Tabela: Invariância dos Descritores.

Descritores de Características Locais					
Detector	Descriptor	Rotação	Escala	Brilho	Ponto de Vista
SIFT	SIFT	✓	✓	✓	✓
SURF	SURF	✓	✓	✓	✓
KAZE	KAZE	✓	✓	✗	✗

Descritores Binários Locais					
Detector	Descriptor	Rotação	Escala	Brilho	Ponto de Vista
-	BRIEF	✗	✗	✓	✗
ORB	ORB	✓	✗	✓	✗
BRISK	BRISK	✓	✓	✓	✓
-	FREAK	✓	✓	✓	✗
AKAZE	AKAZE	✓	✓	✗	✗

Detecção e Correspondência de Características



Figura: Detecção e Correspondência de Características.

Fonte: Elaborado pela autora.

Detecção e Correspondência de Características



Figura: Detecção e Correspondência de Características.
Fonte: Elaborado pela autora.

Saco de Características Visuais

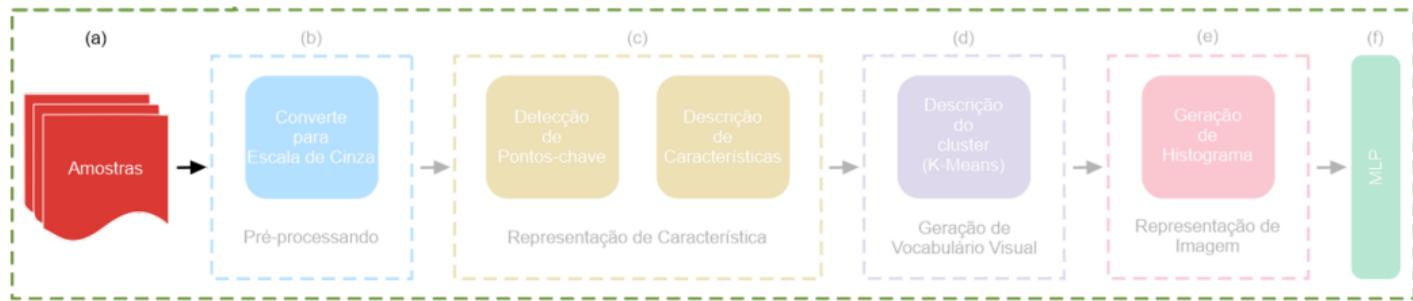


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Saco de Características Visuais

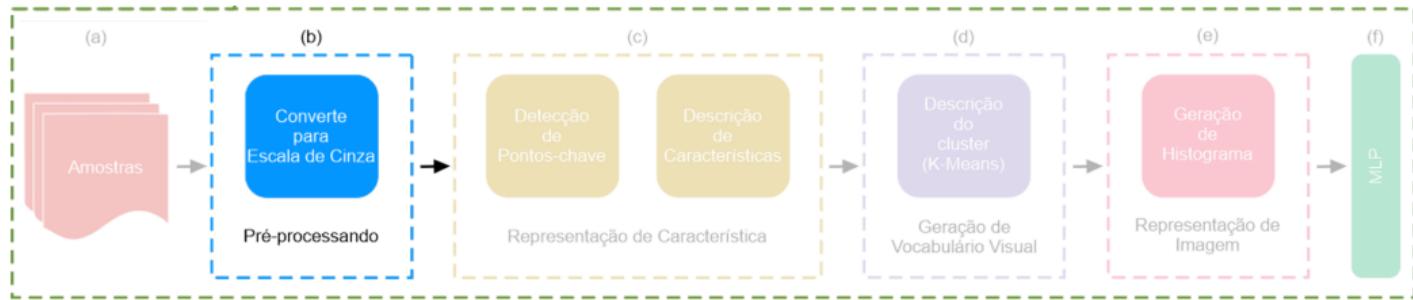


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Saco de Características Visuais

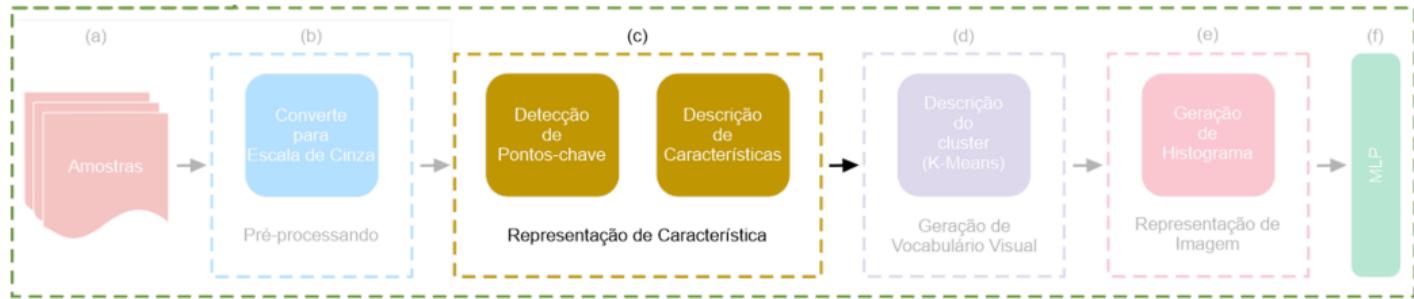


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Saco de Características Visuais

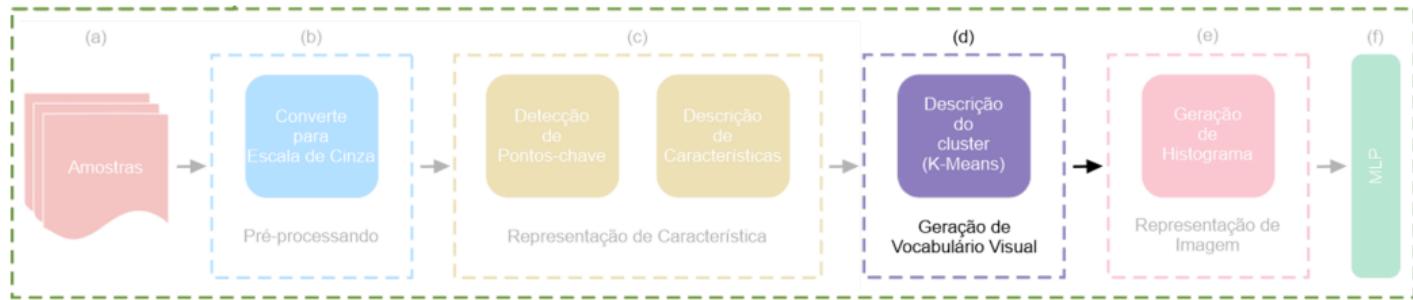


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Saco de Características Visuais

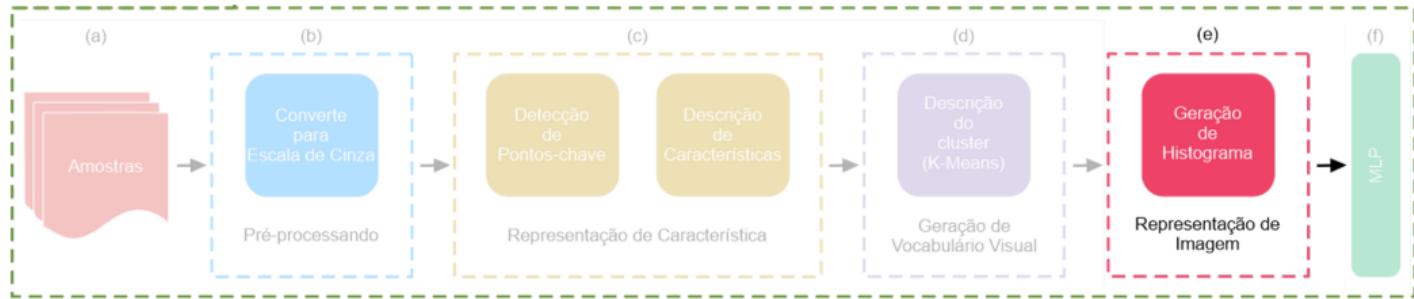


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Saco de Características Visuais

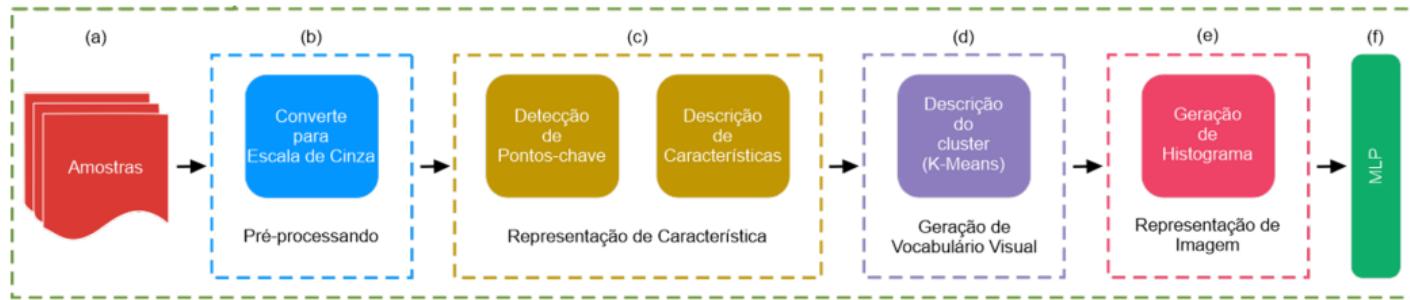


Figura: Abordagem de Saco de Características Visuais com classificador *Perceptron Multicamada*.

Fonte: Elaborado pela autora.

Redes Neurais Convolucionais

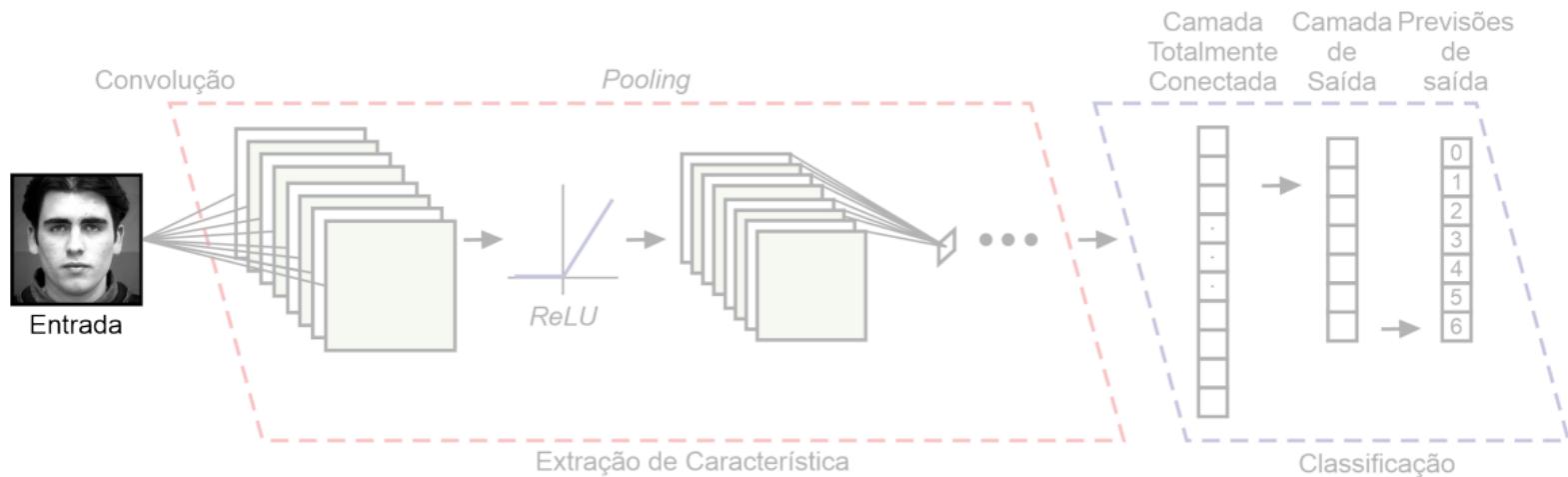


Figura: Arquitetura básica do modelo de Rede Neural Convolucionarial.

Fonte: Adaptado de [3].

Redes Neurais Convolucionais

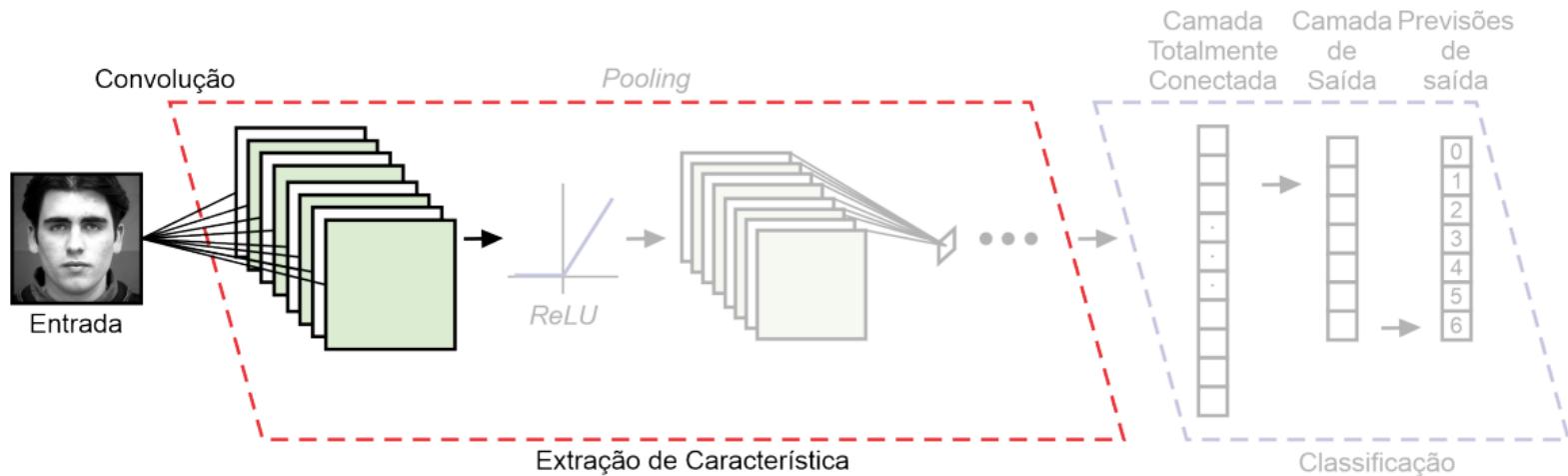


Figura: Arquitetura básica do modelo de Rede Neural Convolucionarial.

Fonte: Adaptado de [3].

Redes Neurais Convolucionais

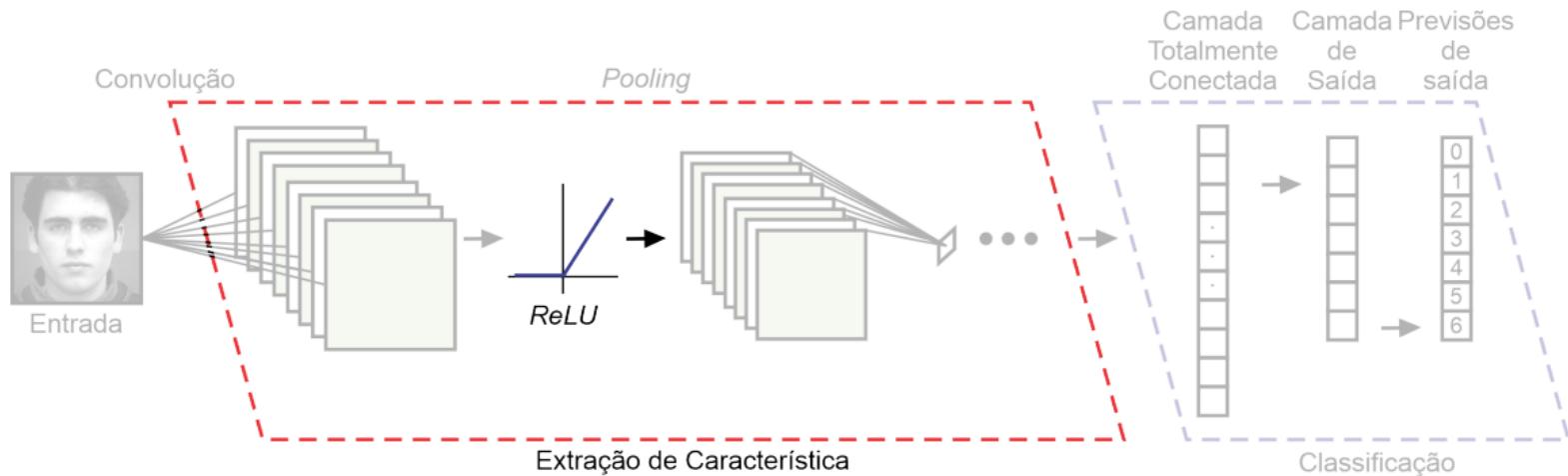


Figura: Arquitetura básica do modelo de Rede Neural Convolucionarial.

Fonte: Adaptado de [3].

Redes Neurais Convolucionais

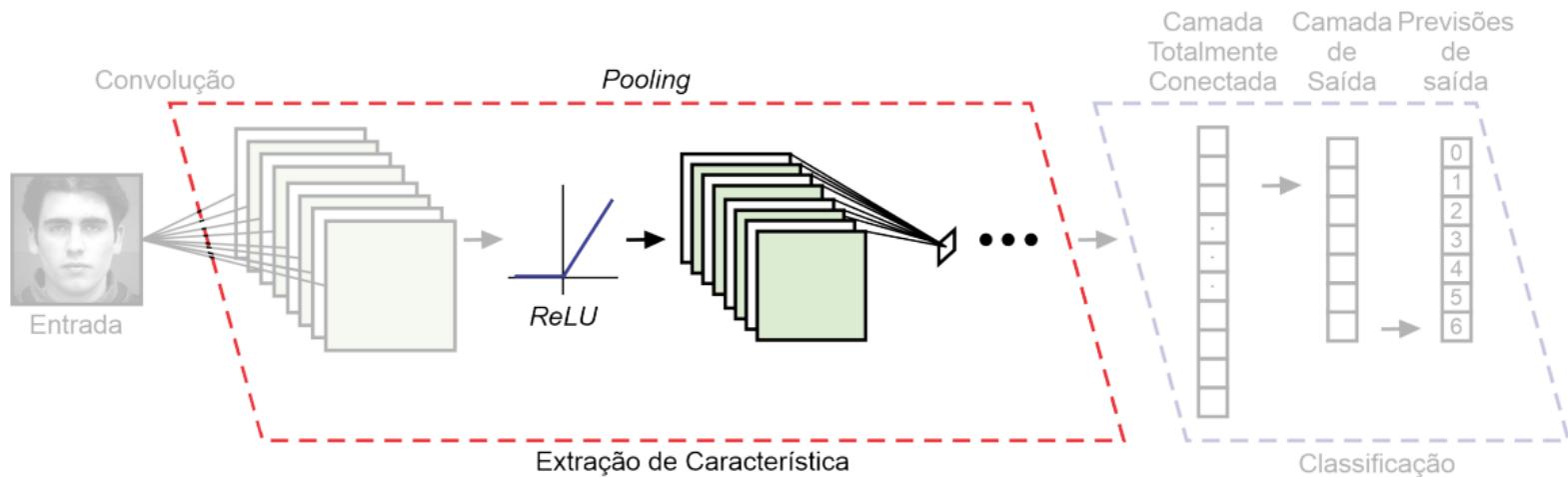


Figura: Arquitetura básica do modelo de Rede Neural Convolucionarial.

Fonte: Adaptado de [3].

Redes Neurais Convolucionais

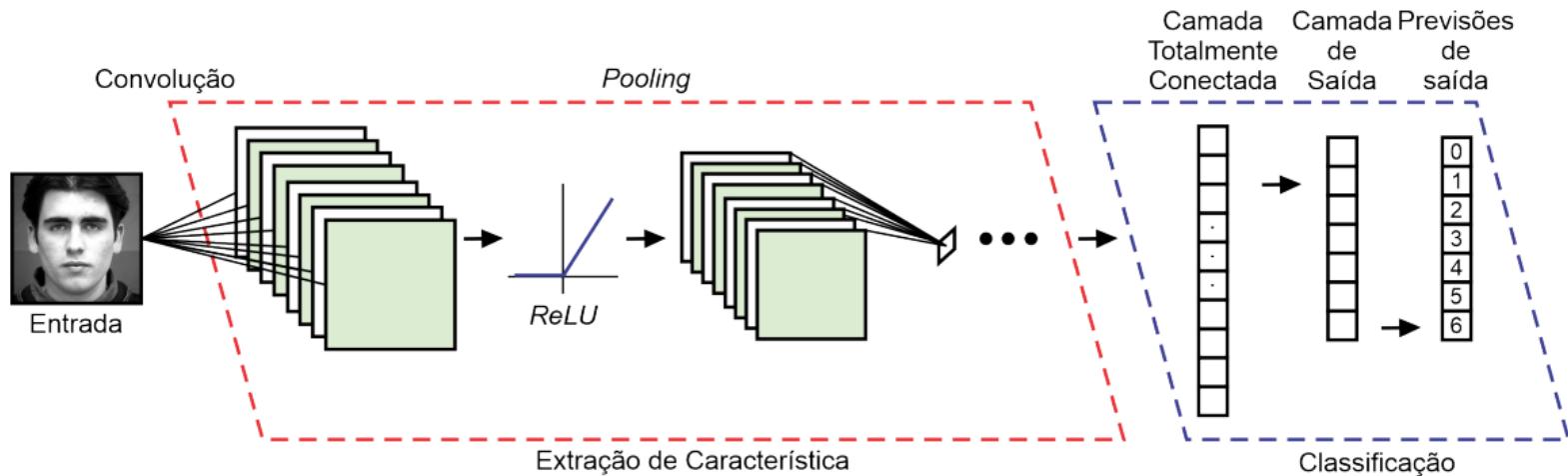


Figura: Arquitetura básica do modelo de Rede Neural Convolucionarial.

Fonte: Adaptado de [3].

Redes Neurais Recorrentes

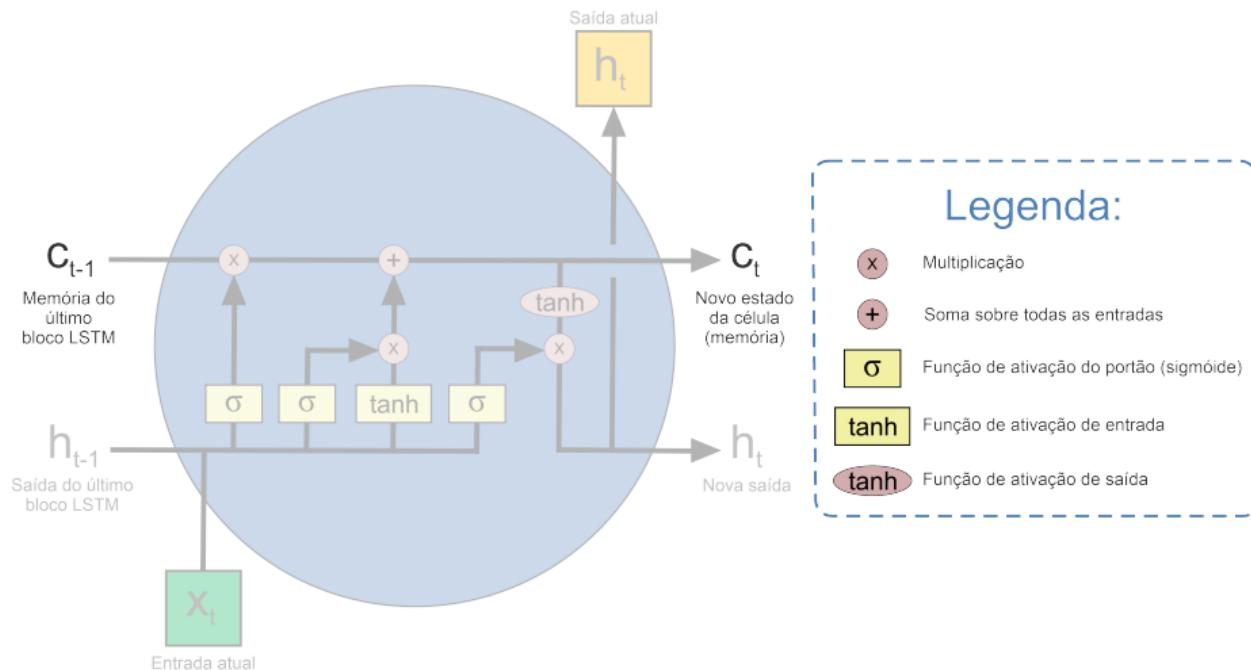


Figura: Bloco do modelo de Rede de Memória Longa de Curto Prazo.

Fonte: Adaptado de [4].

Redes Neurais Recorrentes

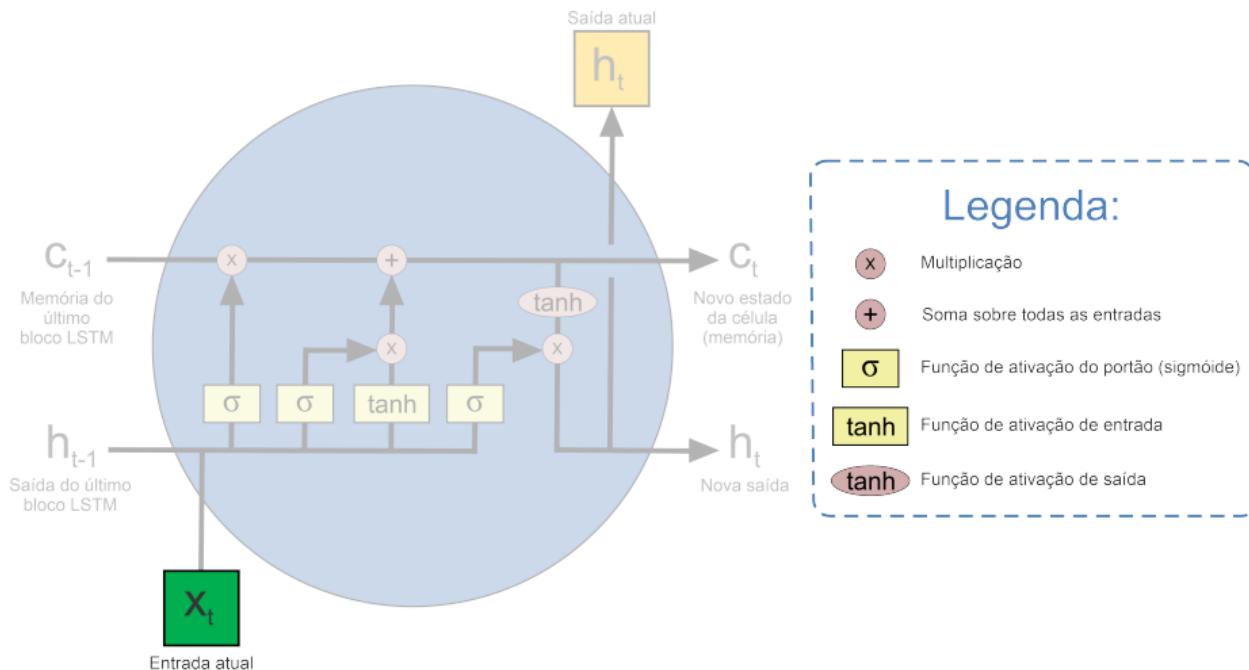


Figura: Bloco do modelo de Rede de Memória Longa de Curto Prazo.

Fonte: Adaptado de [4].

Redes Neurais Recorrentes

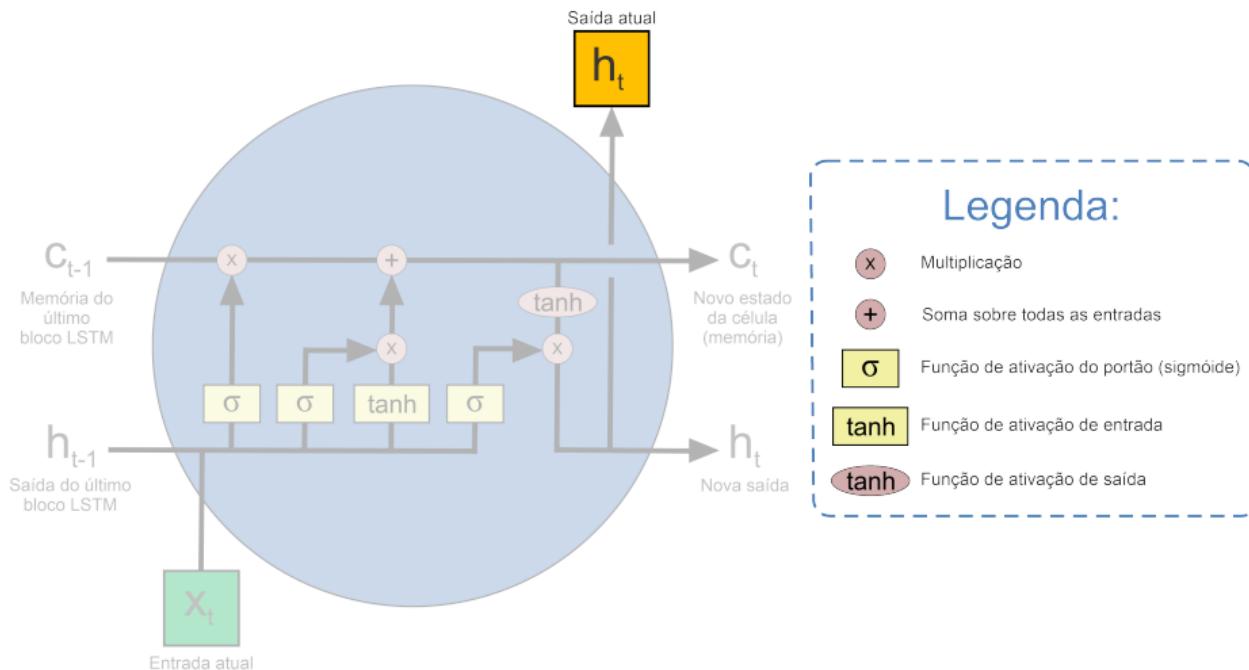


Figura: Bloco do modelo de Rede de Memória Longa de Curto Prazo.

Fonte: Adaptado de [4].

Redes Neurais Recorrentes

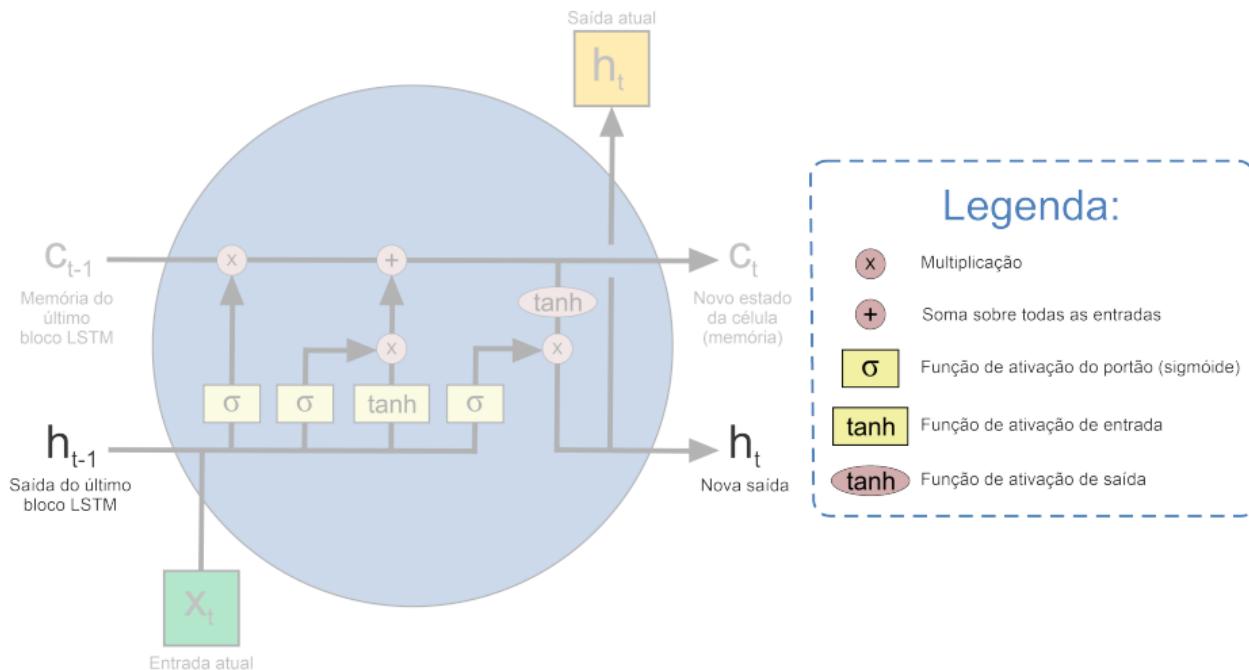


Figura: Bloco do modelo de Rede de Memória Longa de Curto Prazo.

Fonte: Adaptado de [4].

Redes Neurais Recorrentes

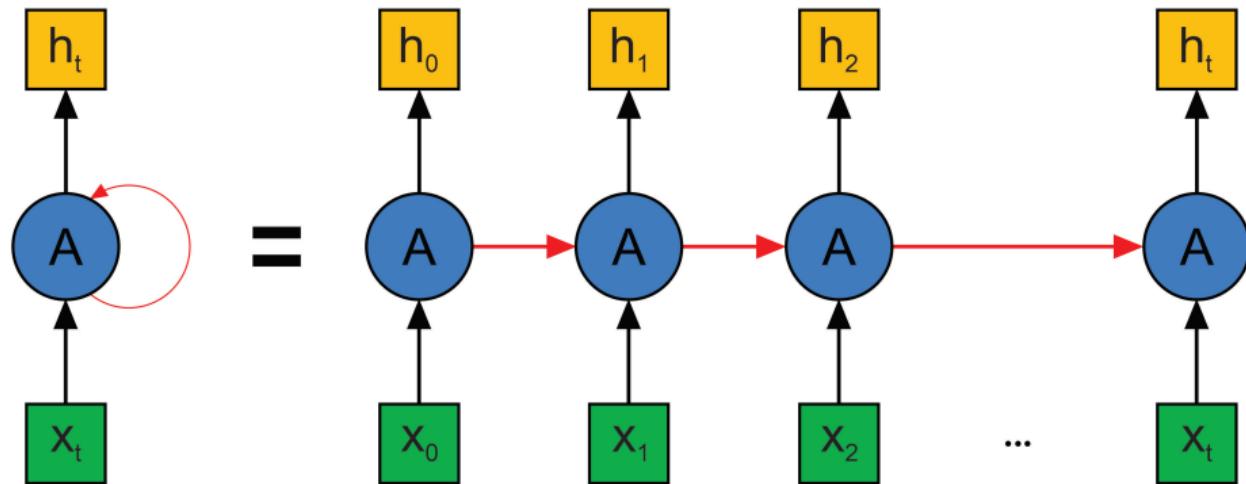


Figura: Representação do modelo de Rede de Memória Longa de Curto Prazo desenrolado.
Fonte: Adaptado de [4].

Modelo Híbrido de Arquitetura de Rede Neural Artificial

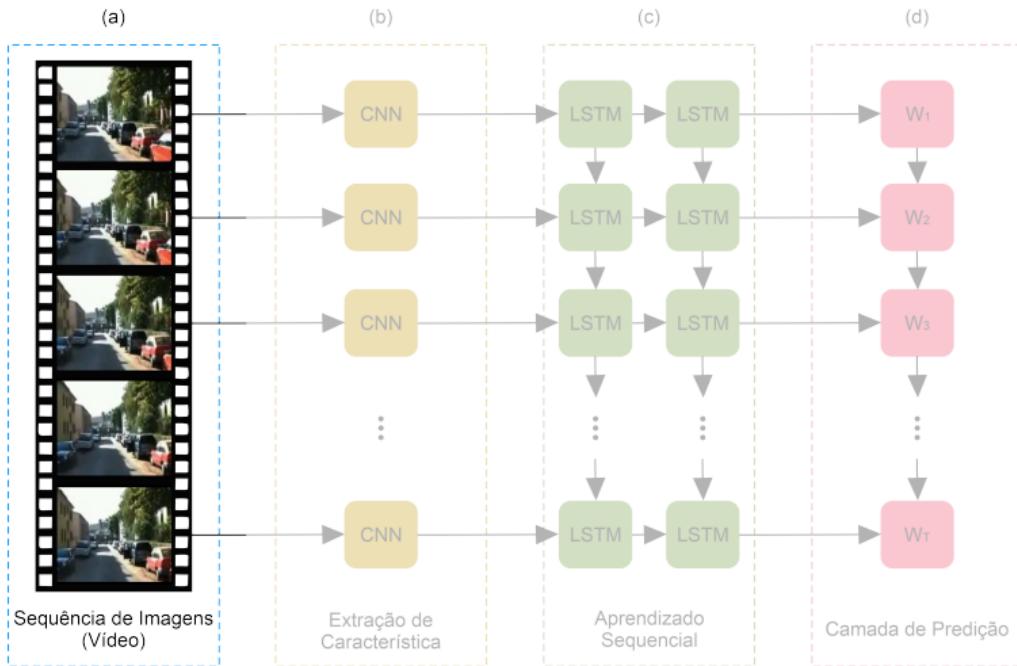


Figura: Arquitetura básica do modelo de Rede Convolucional Recorrente de Longo Prazo.

Fonte: Adaptado de [5].

Modelo Híbrido de Arquitetura de Rede Neural Artificial

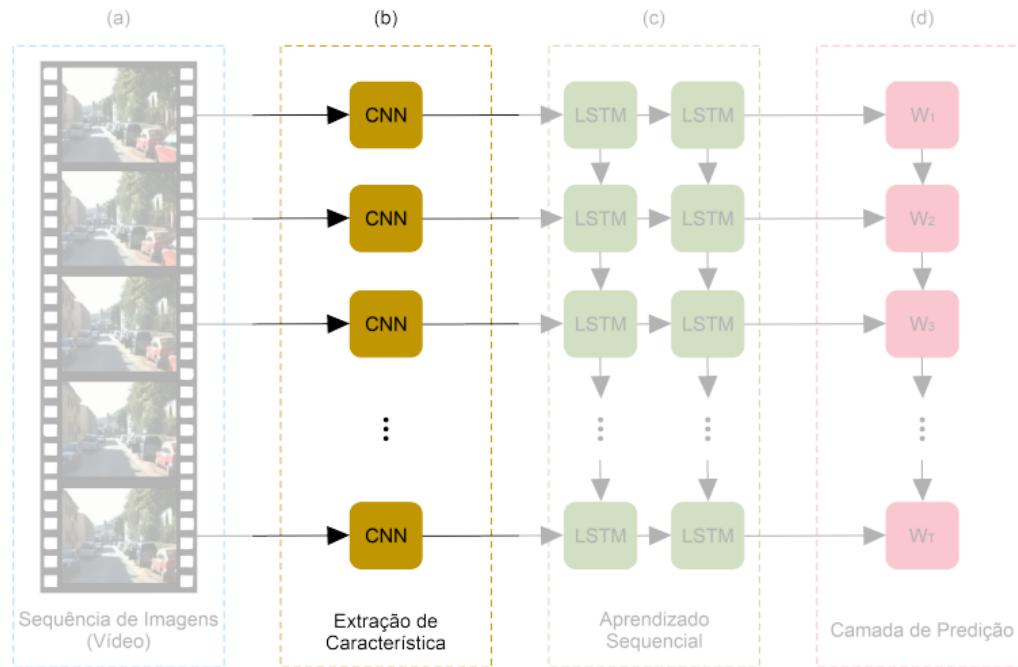


Figura: Arquitetura básica do modelo de Rede Convolucional Recorrente de Longo Prazo.

Fonte: Adaptado de [5].

Modelo Híbrido de Arquitetura de Rede Neural Artificial

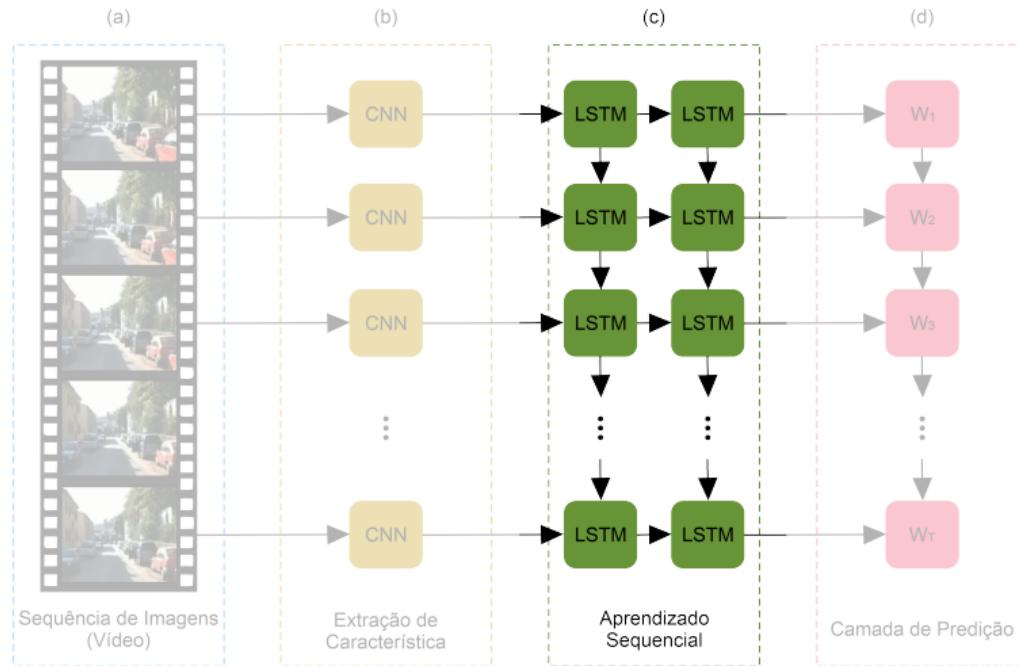


Figura: Arquitetura básica do modelo de Rede Convolucional Recorrente de Longo Prazo.

Fonte: Adaptado de [5].

Modelo Híbrido de Arquitetura de Rede Neural Artificial

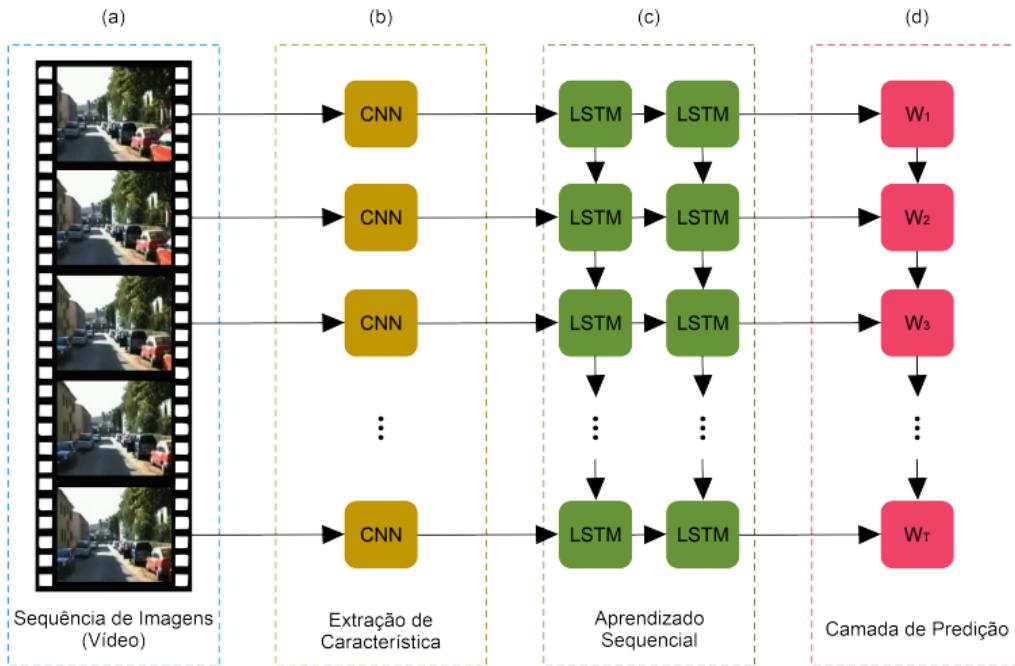


Figura: Arquitetura básica do modelo de Rede Convolucional Recorrente de Longo Prazo.

Fonte: Adaptado de [5].

Comparativo entre Trabalhos Relacionados e Abordagem Proposta

Tabela: Comparação entre métodos utilizados pelos principais trabalhos relacionados.

Métodos	Trabalhos											
	ORB-PCA [6]	TEX-Nets [7]	LF-Net [8]	Key-Net [9]	LBP Dual-Channel CNN [10]	LBCNN [11]	BinaryConnect [12]	Binary-Weight e XNOR [13]	BNIN-Binary-Weight [14]	SCNN [15]	LCD OverFeat-based [16]	Abordagem Proposta
Aprendizado Profundo	x	✓	x	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pesos Binários	x	x	x	x	x	✓	✓	✓	✓	x	x	✓
Pesos Esparsos	x	x	x	x	x	✓	x	x	x	✓	x	✓
Extração de Características baseada em Descritores	✓	✓	✓	✓	✓	✓	x	x	x	x	✓	
Redução de Dimensionalidade	✓	x	x	x	x	x	x	x	x	x	✓	
Detecção de Fechamento de Loop	x	x	x	x	x	x	x	x	x	✓	✓	

DescNet — Rede Neural Convolucional de Descritores

Camadas:

- Detecção de Característica Local;
- Convolução de Descriptor Local.

Arquitetura básica de Rede Neural Convolucional de Descritores

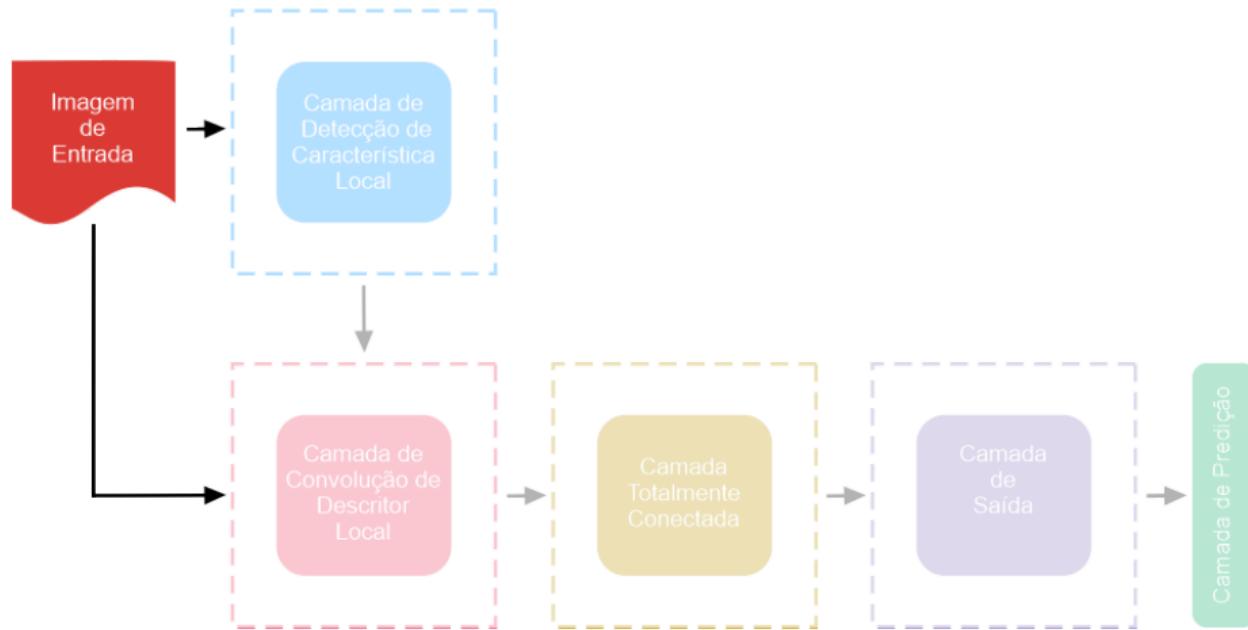


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Arquitetura básica de Rede Neural Convolucional de Descritores

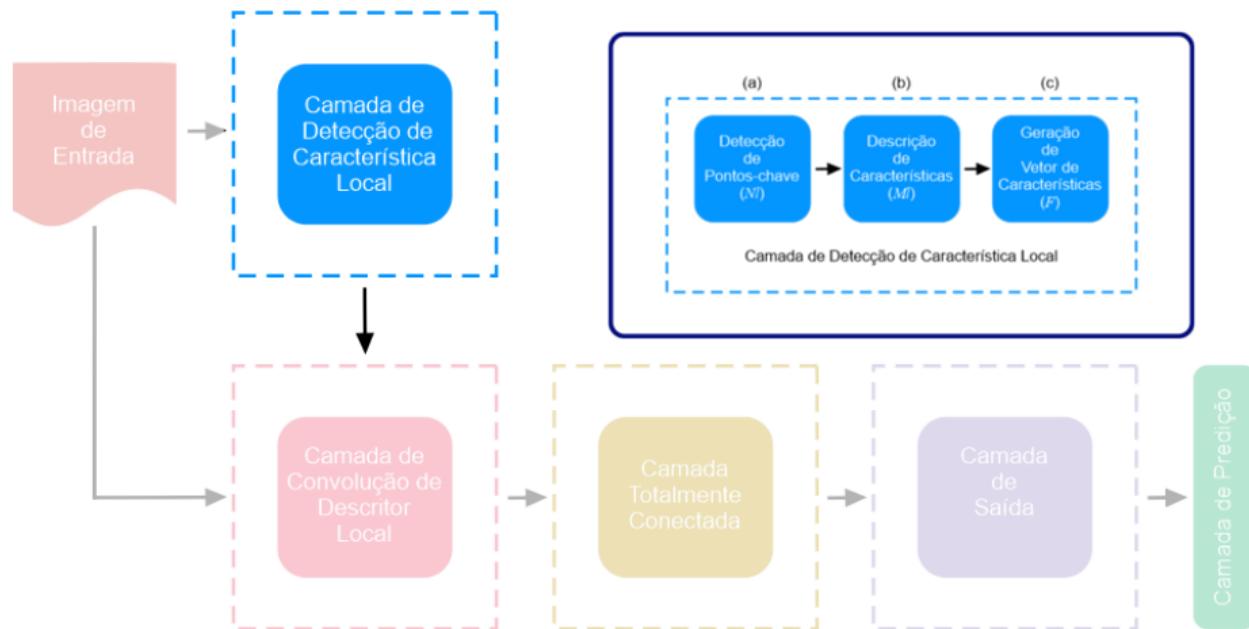


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Arquitetura básica de Rede Neural Convolucional de Descritores

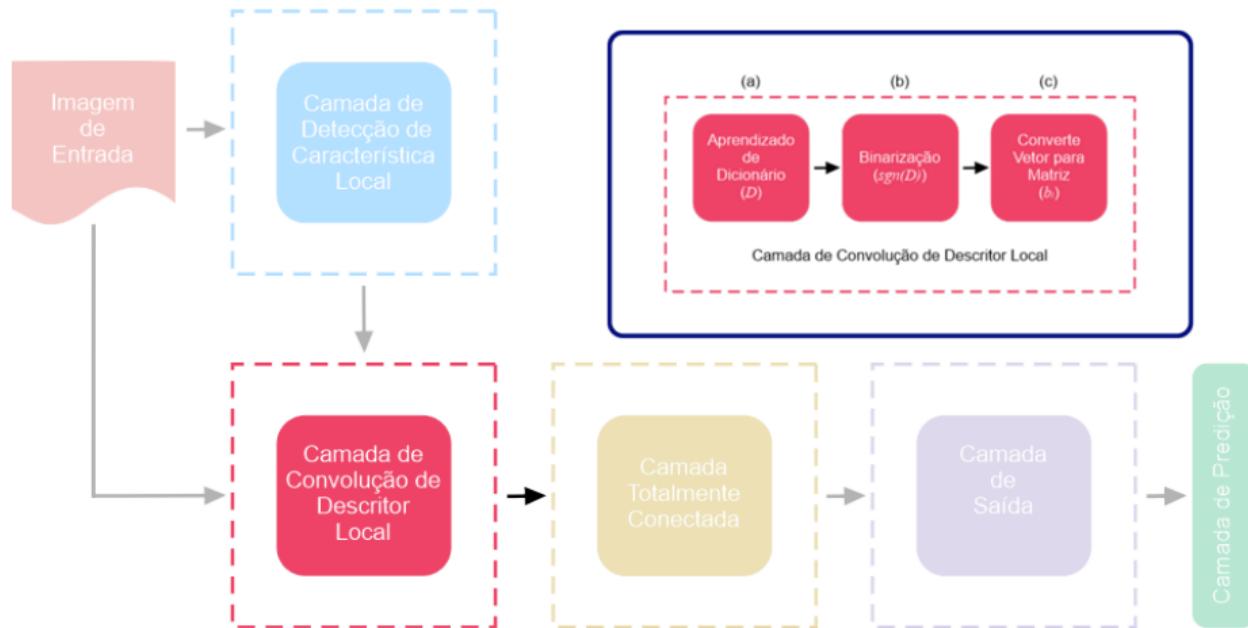


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Onde, b_i são os pesos de descritor, gerados a partir da matriz de transformação de $sgn(D)$.

Comportamento da operação de convolução

Comportamento da operação de convolução

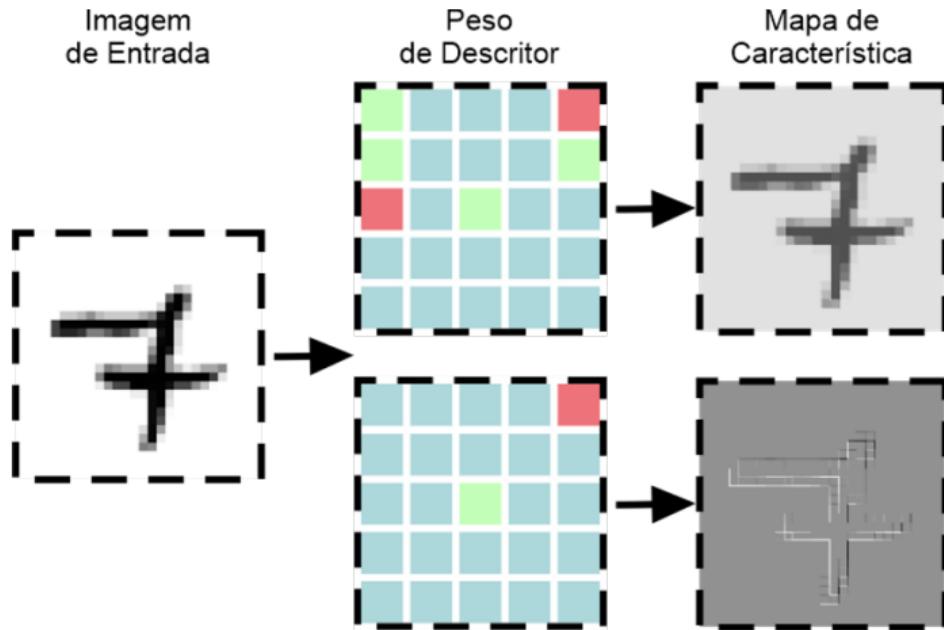


Figura: Comportamento da operação de convolução.

Fonte: Elaborado pela autora.

Arquitetura básica de Rede Neural Convolucional de Descritores

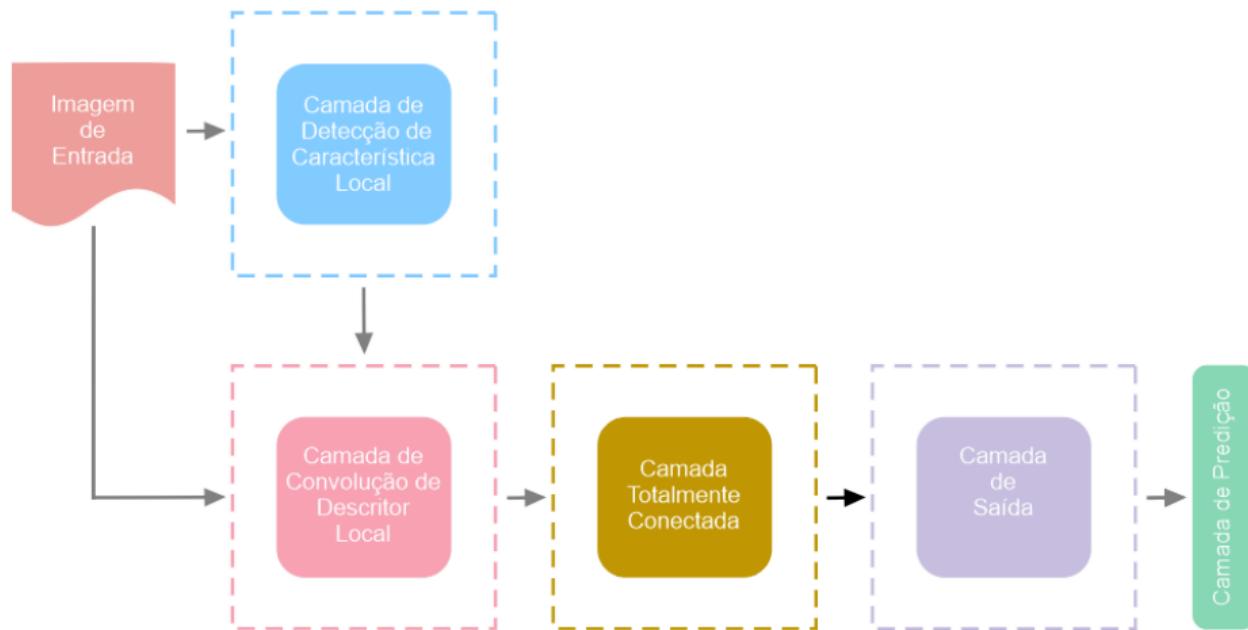


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Arquitetura básica de Rede Neural Convolucional de Descritores

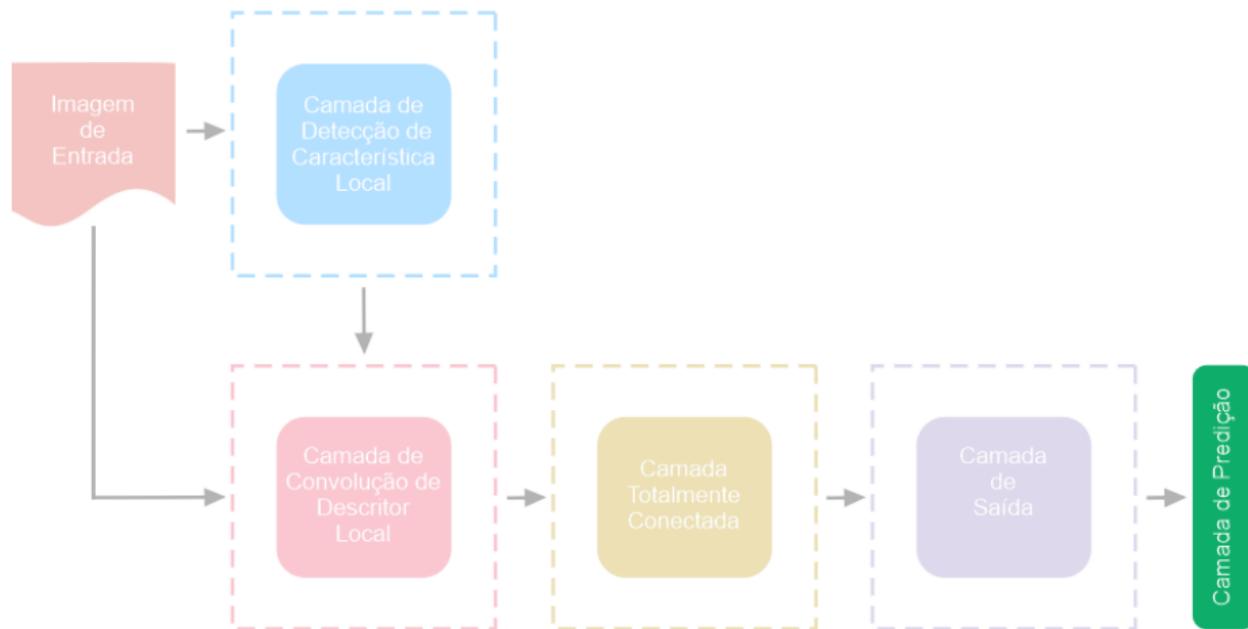


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Arquitetura básica de Rede Neural Convolucional de Descritores

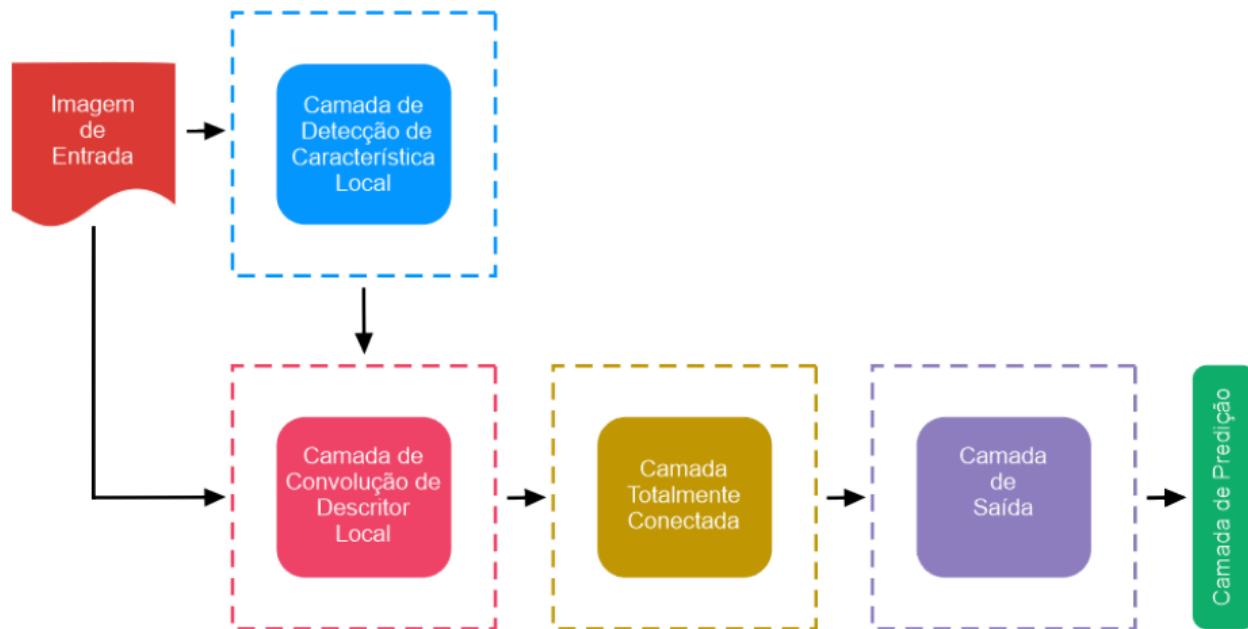


Figura: Arquitetura básica do modelo de Rede Neural Convolucional de Descritores.
Fonte: Elaborado pela autora.

Kit de Desenvolvimento NVIDIA Jetson Nano

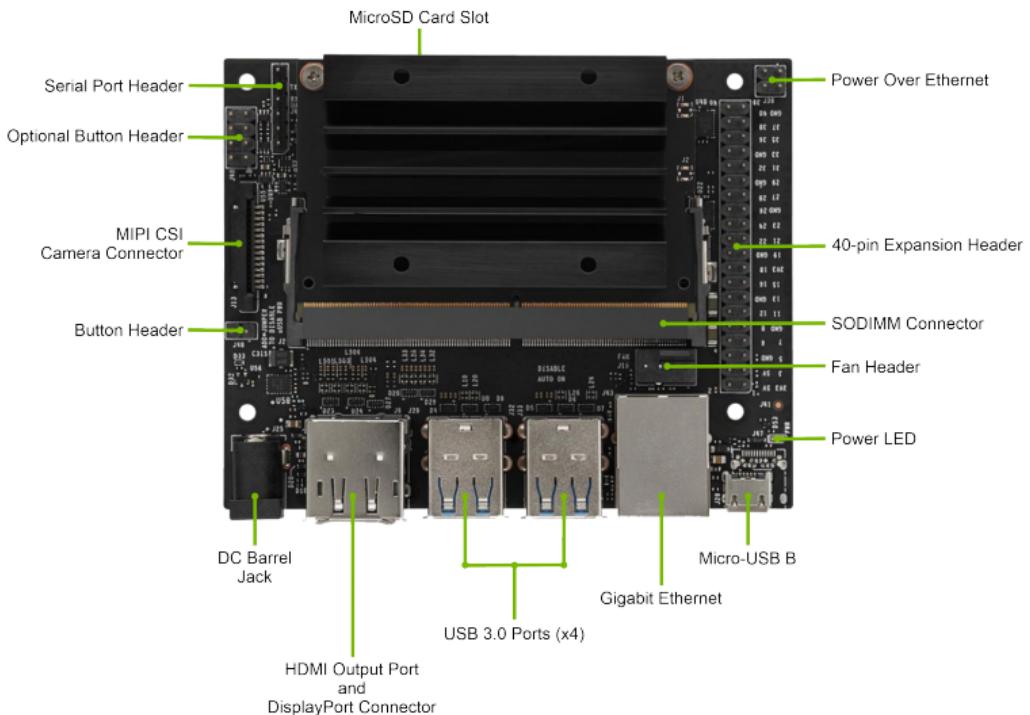


Figura: Visão esquemática do microcomputador NVIDIA Jetson Nano.

Avaliação comparativa entre Descritores

Conjuntos de dados visuais:

- MNIST
- JAFFE
- Extended CK+
- FEI
- CIFAR-10
- FER-2013

Avaliação comparativa entre Descritores

Descritores Binários Locais:

- BRIEF
- ORB
- BRISK
- FREAK
- AKAZE

Descritores de Recursos Locais:

- SIFT
- SURF
- KAZE

Etapa de Treinamento

Tabela: Tempo de processamento (min.) na etapa de Representação de Características da abordagem de Saco de Características Visuais da etapa de treinamento.

Etapa de Representação de Características					
Algoritmos	Dataset	Conjunto de dados visuais			
		MNIST	Extended CK+	CIFAR-10	FER-2013
SIFT	Treinamento	02:37	01:01	02:59	02:51
SURF		00:49	00:16	00:54	00:25
KAZE		02:59	01:02	02:59	02:52
ORB		00:32	00:17	00:46	00:39
BRISK		-	00:29	00:57	01:00

Etapa de Treinamento

Tabela: Tempo de processamento (min.) na etapa de Geração de Vocabulário Visual da abordagem de Saco de Características Visuais da etapa de treinamento.

Geração de Vocabulário Visual					
Algoritmos	Dataset	Conjunto de dados visuais			
		MNIST	Extended CK+	CIFAR-10	FER-2013
SIFT	Treinamento	02:30	00:48	05:14	02:50
SURF		00:01	00:05	00:03	00:20
KAZE		01:45	00:39	05:09	04:13
ORB		02:39	01:46	03:22	02:43
BRISK		-	00:06	00:00	00:06

Etapa de Treinamento

Tabela: Tempo de processamento (min.) na etapa de Representação de Imagem da abordagem de Saco de Características Visuais da etapa de treinamento.

Etapa de Representação de Imagem					
Algoritmos	Dataset	Conjunto de dados visuais			
		MNIST	Extended CK+	CIFAR-10	FER-2013
SIFT	Treinamento	01:04	00:11	00:56	00:33
SURF		00:13	00:10	00:20	00:29
KAZE		01:04	00:11	00:56	00:33
ORB		01:02	00:12	00:55	00:33
BRISK		-	00:10	00:01	00:19

Etapa de Teste

Tabela: Tempo de processamento (min.) nas etapas de Representação de Características e Representação de Imagem da abordagem de Saco de Características Visuais da etapa de teste.

		Etapa de Representação de Características e Representação de Imagem			
Algoritmos	Dataset	Conjunto de dados visuais			
		MNIST	Extended CK+	CIFAR-10	FER-2013
SIFT	Teste	00:35	00:04	00:46	00:25
SURF		00:09	00:01	00:08	00:06
KAZE		00:39	00:04	00:46	00:25
ORB		00:16	00:01	00:20	00:09
BRISK		-	00:02	00:11	00:09

Etapa de Teste

Tabela: Taxa de acurácia (%) nas etapas de teste dos descritores BRIEF, AKAZE e FREAK no conjunto de dados visuais FEI.

Algoritmos	Dataset	Modelo Perceptron Multicamadas					
		MLP1	MLP2	MLP3	MLP4	MLP5	MLP6
BRIEF	FEI	0.78	0.74	0.77	0.76	0.82	0.85
AKAZE		0.85	0.87	0.83	0.84	0.83	0.86
FREAK		0.47	0.47	0.47	0.51	0.51	0.54

Etapa de Teste

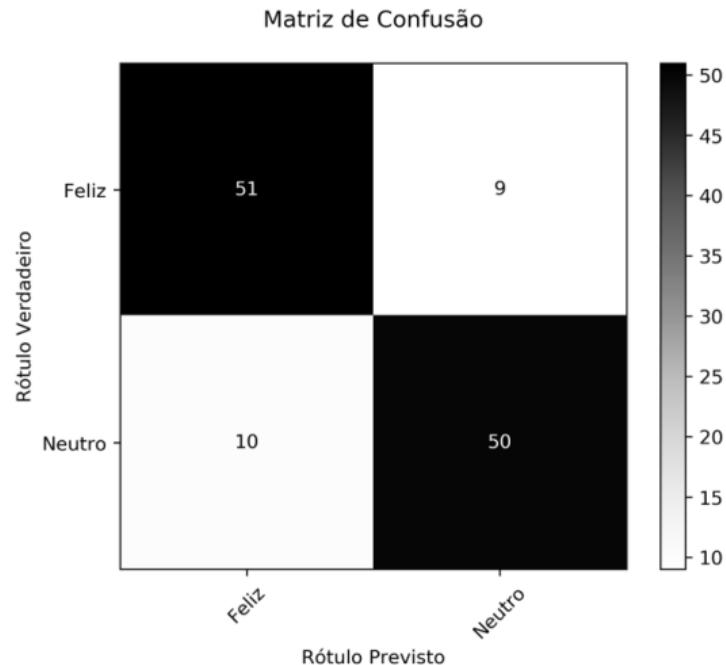


Figura: Matriz de Confusão sobre conjunto de dados visuais FEI com o descritor BRISK com o modelo MLP5.
Fonte: Elaborado pela autora.

Detalhes da Implementação

- **Linha de base:** ResNet-132;
- **Descriptor:** BRISK;
- **Função de ativação:** ReLU;
- **Algoritmos de otimização:** Adam;
- **Taxa de aprendizado:** $1e-3$;
- **Taxas de Aprendizado Cílicos:** Intervalo entre $1e-8$ e $1e0$ e 10 épocas.

Detalhes da Implementação



Figura: Teste de Variação da Taxa de Aprendizado.
Fonte: Elaborado pela autora.

Resultados de treinamento no conjunto de dados visuais MNIST

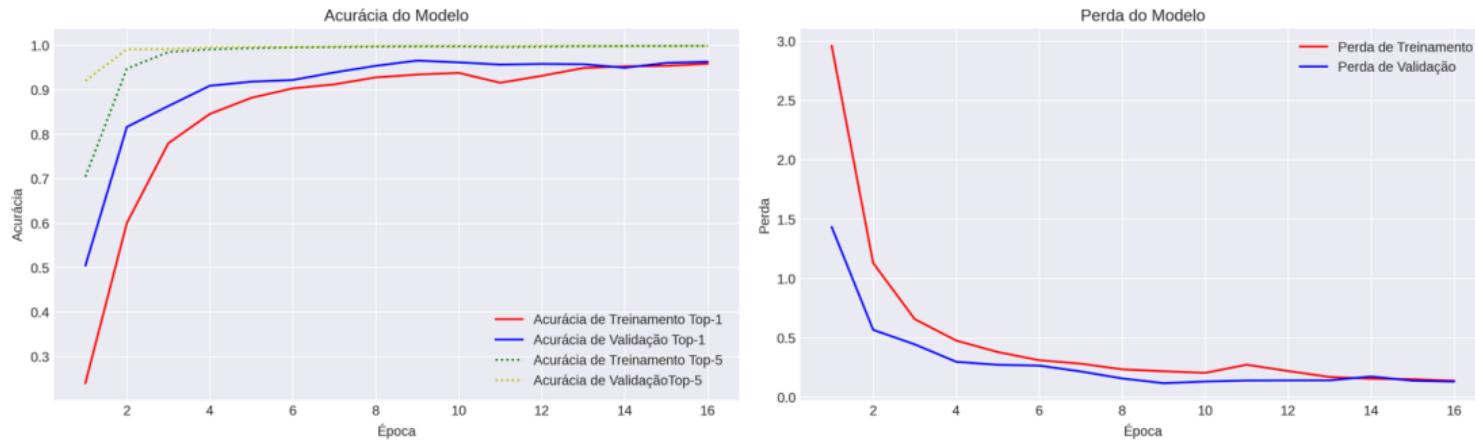


Figura: Curvas de aprendizado no conjunto de dados visuais MNIST.

Fonte: Elaborado pela autora.

Resultados de treinamento no conjunto de dados visuais CIFAR-10

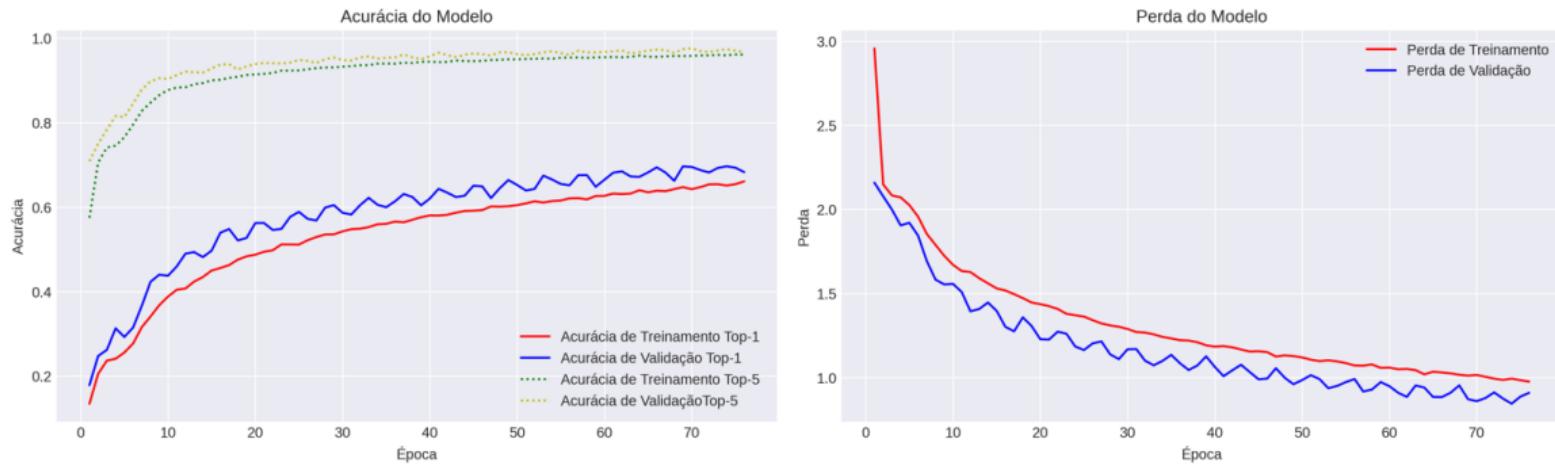


Figura: Curvas de aprendizado no conjunto de dados visuais CIFAR-10.

Fonte: Elaborado pela autora.

Resultados de teste e Comparação entre DescNet e Linha de Base

Tabela: Resultados de treinamento e Comparação entre DescNet e Linha de Base.

Conjunto de dados visuais	DescNet				Linha de Base			
	Top-1	Top-5	Parâmetros aprendíveis	Tempo de Treinamento	Top-1	Top-5	Parâmetros aprendíveis	Tempo de Treinamento
MNIST	56%	90%	50,307,978	01:01:02	67%	89%	58,295,178	02:59:54
CIFAR-10	23%	65%	50,307,978	05:29:46	38%	77%	58,295,178	07:44:53

Conclusão

A proposta da DescNet em dois conjuntos de dados visuais (MNIST e CIFAR-10) obteve resultados significativos, ao mesmo tempo, em que permite economia no número de parâmetros do modelo e, consequentemente, economia computacional significativa. Tornando a DescNet um modelo aplicável em ambientes reais com recursos escassos e limitados.

Nossa abordagem é promissora, onde os métodos propostos possuem o potencial de realizar a tarefa de detecção de fechamento de loop de um sistema VSLAM eficaz e com desempenho eficiente em trabalhos futuros.

Trabalhos Publicados

- A. Raibolt, A. Angonese and P. Rosa, **Comparative Evaluation of Feature Descriptors Through Bag of Visual Features with Multilayer Perceptron on Embedded GPU System** 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE), 2020, pp. 1-6, doi: 10.1109/LARS/SBR/WRE51543.2020.9306931.

Artigo escolhido entre os 3 melhores artigos publicados no Simpósio Latino-Americano de Robótica/Simpósio Brasileiro de Robótica (LARS/SBR 2020) em novembro de 2020;

Trabalhos Publicados

- A. Raibolt, A. Angonese, G. Alves, P. Rosa, **Towards loop Closure Detection for SLAM Applications using Bag of Visual Features: Experiments and Simulation** Computational Neuroscience: Third Latin American Workshop, LWCN 2021. (aceito para publicação nos Anais da Conferência);
- A. Raibolt, A. Angonese, G. Alves, P. Rosa, **Descriptor Convolutional Neural Network (DescNet)**. (a ser submetido em Periódico).

Repositórios *git*

- <https://github.com/whoisraibolt/Feature-Detection-and-Description>
- <https://github.com/whoisraibolt/Feature-Detection-and-Matching>
- <https://github.com/whoisraibolt/BoVF-with-MLP-Classifier>
- <https://github.com/whoisraibolt/Reform-Conv-Filters-through-Descriptors>
- <https://github.com/whoisraibolt/DescNet>

 52  15
Stars Forks

- <https://dzone.com/articles/run-hundreds-of-experiments-with-opencv-and-hydra>

Referências I

- [1] V. C. Guizilini, *Localização e mapeamento simultâneos com auxílio visual omnidirecional.* PhD thesis, Universidade de São Paulo, 2008.
- [2] S. Thrun, W. Burgard, and D. Fox, "Probabilistic robotics. 2005," *Massachusetts Institute of Technology, USA*, 2005.
- [3] M. Peemen, B. Mesman, and H. Corporaal, "Efficiency optimization of trainable feature extractors for a consumer platform," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 293–304, Springer, 2011.
- [4] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.
- [5] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2625–2634, 2015.
- [6] A. Vinay, C. A. Kumar, G. R. Shenoy, K. N. B. Murthy, and S. Natarajan, "Orb-pca based feature extraction technique for face recognition," *Procedia Computer Science*, vol. 58, pp. 614–621, 2015.
Second International Symposium on Computer Vision and the Internet (VisionNet'15).
- [7] R. M. Anwer, F. S. Khan, J. van de Weijer, M. Molinier, and J. Laaksonen, "Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification," *ISPRS journal of photogrammetry and remote sensing*, vol. 138, pp. 74–85, 2018.

Referências II

- [8] Y. Ono, E. Trulls, P. Fua, and K. M. Yi, "Lf-net: Learning local features from images," *arXiv preprint arXiv:1805.09662*, 2018.
- [9] A. Barroso-Laguna, E. Riba, D. Ponsa, and K. Mikolajczyk, "Key. net: Keypoint detection by handcrafted and learned cnn filters," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5836–5844, 2019.
- [10] X. Wei, X. Yu, B. Liu, and L. Zhi, "Convolutional neural networks and local binary patterns for hyperspectral image classification," *European Journal of Remote Sensing*, vol. 52, no. 1, pp. 448–462, 2019.
- [11] F. Juefei-Xu, V. Naresh Boddeti, and M. Savvides, "Local binary convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 19–28, 2017.
- [12] M. Courbariaux, Y. Bengio, and J.-P. David, "Binaryconnect: Training deep neural networks with binary weights during propagations," *Advances in neural information processing systems*, vol. 28, pp. 3123–3131, 2015.
- [13] M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, "Xnor-net: Imagenet classification using binary convolutional neural networks," in *European conference on computer vision*, pp. 525–542, Springer, 2016.
- [14] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Binarized neural networks," *Advances in neural information processing systems*, vol. 29, pp. 4107–4115, 2016.
- [15] B. Liu, M. Wang, H. Foroosh, M. Tappen, and M. Pensky, "Sparse convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 806–814, 2015.

Referências III

- [16] X. Zhang, Y. Su, and X. Zhu, "Loop closure detection for visual slam systems using convolutional neural network," in *2017 23rd International Conference on Automation and Computing (ICAC)*, pp. 1–6, IEEE, 2017.

INSTITUTO MILITAR DE ENGENHARIA - IME
PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS E COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

DescNet — Rede Neural Convolucional de Descritores: uma abordagem para o fechamento de *loop* em *Visual SLAM*



Alexandra Miguel Raibolt da Silva

Paulo Fernando Ferreira Rosa
(Orientador)

Alberto Torres Angonese
(Co-orientador)

