

DESAFIOS NA CONSTRUÇÃO DE UM SISTEMA DE NAVEGAÇÃO DE ROBÔS AUTÔNOMOS MÓVEIS BASEADO EM VISÃO: DETECÇÃO DE FECHAMENTO DE LOOP

Alexandra Miguel Raibolt da Silva
(raibolt@ime.eb.br)



Sumário

- Introdução
- Conceitos Básicos e Estado da Arte
- A Proposta
- Viabilidade
- Experimentos e Resultados
- Conclusão
- Referências

Introdução

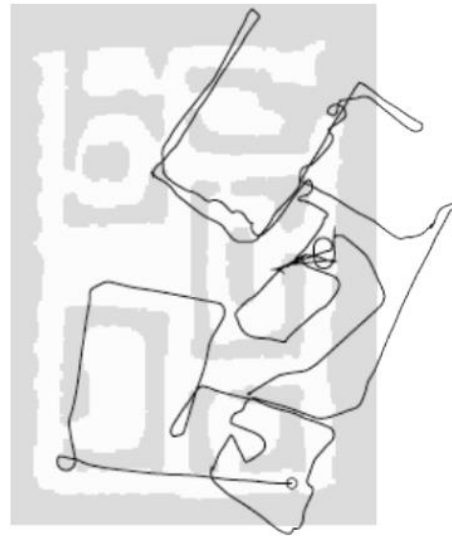
O problema de SLAM

- Pode ser caracterizado pela resolução de tarefas complexas, tais quais:
 - Mapeamento;
 - Localização.
- Estas duas tarefas não podem ser realizadas de formas independentes, isto é, elas atuam de forma complementar.

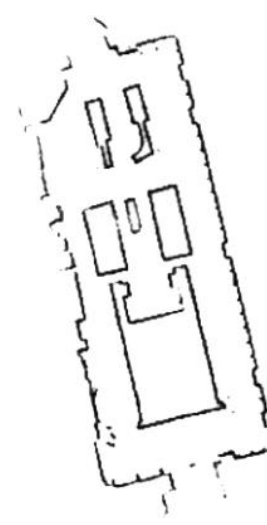
Influência da incerteza na localização e mapeamento



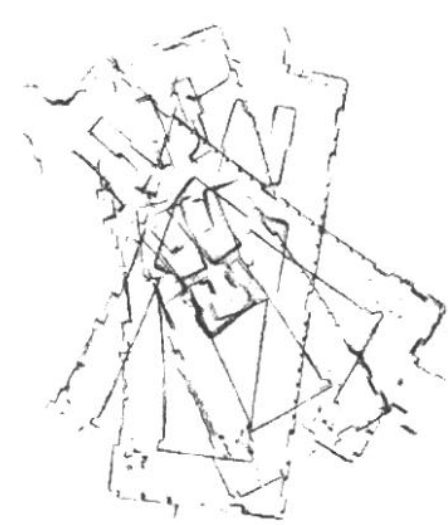
Localização
sem erros



Localização
com erros



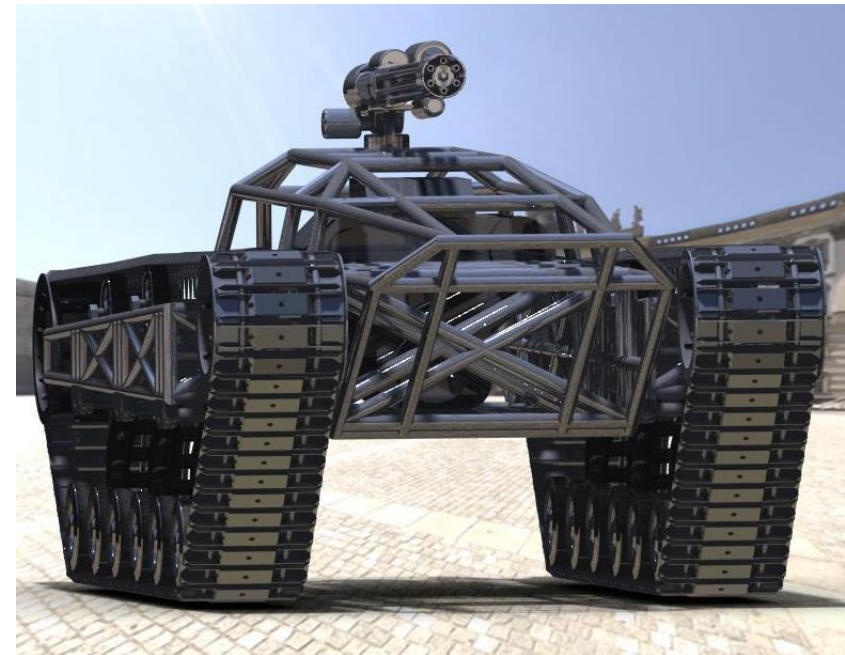
Mapeamento
sem erros



Mapeamento
com erros

Veículos Terrestres Não-Tripulados (VTNTs)

- Aplicações do uso de VTNTs:
 - Defesa;
 - Operações militares;
 - Busca;
 - Resgate e salvamento;
 - Serviços domésticos.



Odometria Visual + SLAM = VSLAM

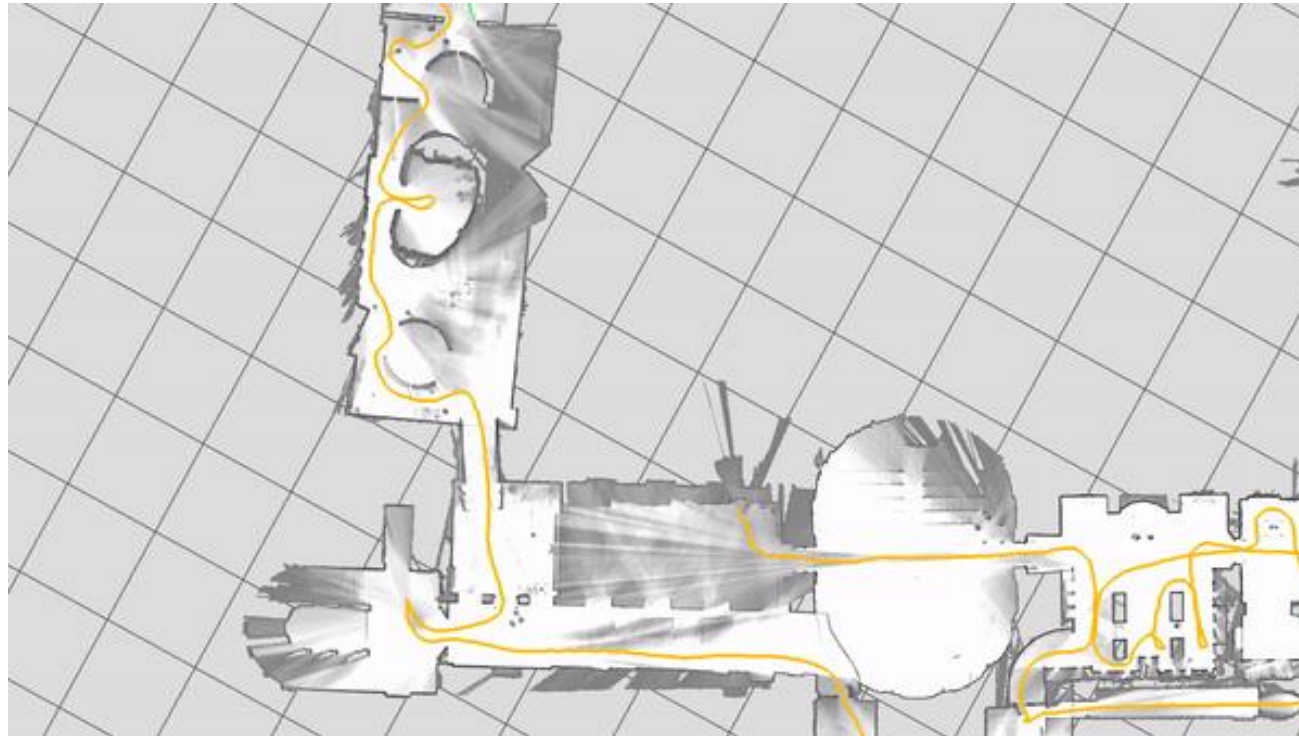
- Vantagens VSLAM:

- 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=-EQAJOoRqEQ>

Caracterização do Problema

- Técnicas de Aprendizado de Máquina baseadas em Aprendizado Profundo, como por exemplo, Redes Neurais Convolucionais, apesar de serem exploradas em sistemas VSLAM, são pouco exploradas para a tarefa de Detecção de Fechamento de Loop.
- Entretanto, um problema ainda enfrentado ao treinar tais arquiteturas, está relacionado ao poder computacional necessário, onde, torna-se um recurso caro, ou até mesmo indisponível.

Motivação

- Buscamos alternativas baratas, robustas e com desempenho computacional eficiente.
- Descartando a utilização de recursos como sensores de distância e lasers, que, comumente são recursos caros e que demandam de um poder computacional alto.

Objetivo

- Integração de um modelo híbrido de arquitetura de Rede Neural Artificial com um sistema de Localização e Mapeamento Simultâneos baseado em visão capaz de solucionar o problema de Detecção de Fechamento de Loop.

Conceitos Básicos e Estado da Arte

Odometria Visual



Câmera
Monocular



Câmera
Stereo



Câmera
Omnidirecional



Câmera
RGB-D

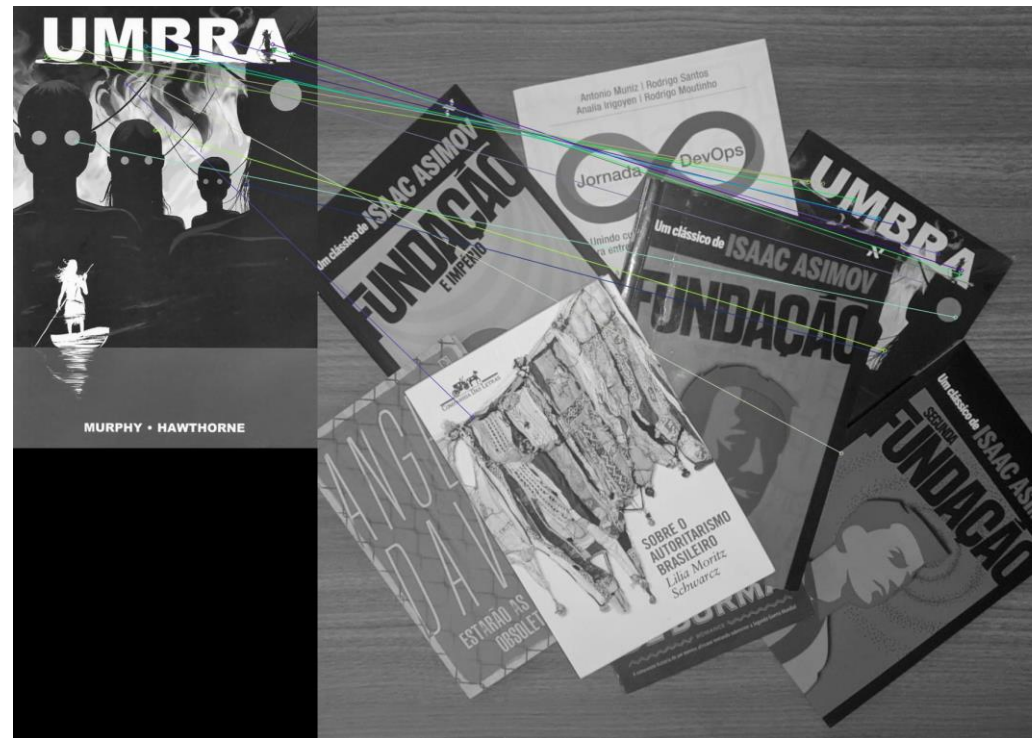
Detecção e Descrição de Características

- Um Descritor de Recurso Local ou um Descritor Binário Local exerce a função de encontrar características “interessantes” presentes em uma imagem.
- Não importa como a imagem é modificada (e.g., rotação, escala, redução, expansão, distorção, etc.), sempre encontraremos as mesmas características para esta mesma imagem.

Detecção e Descrição de Características

Detector	Descritor	Rotação	Escala	Brilho	Ponto de Vista
SIFT	SIFT	✓	✓	✓	✓
SURF	SURF	✓	✓	✓	✓
KAZE	KAZE	✓	✓	x	x
-	BRIEF	x	x	✓	x
ORB	ORB	✓	x	✓	x
BRISK	BRISK	✓	✓	✓	✓
-	FREAK	✓	✓	✓	x
AKAZE	AKAZE	✓	✓	x	x

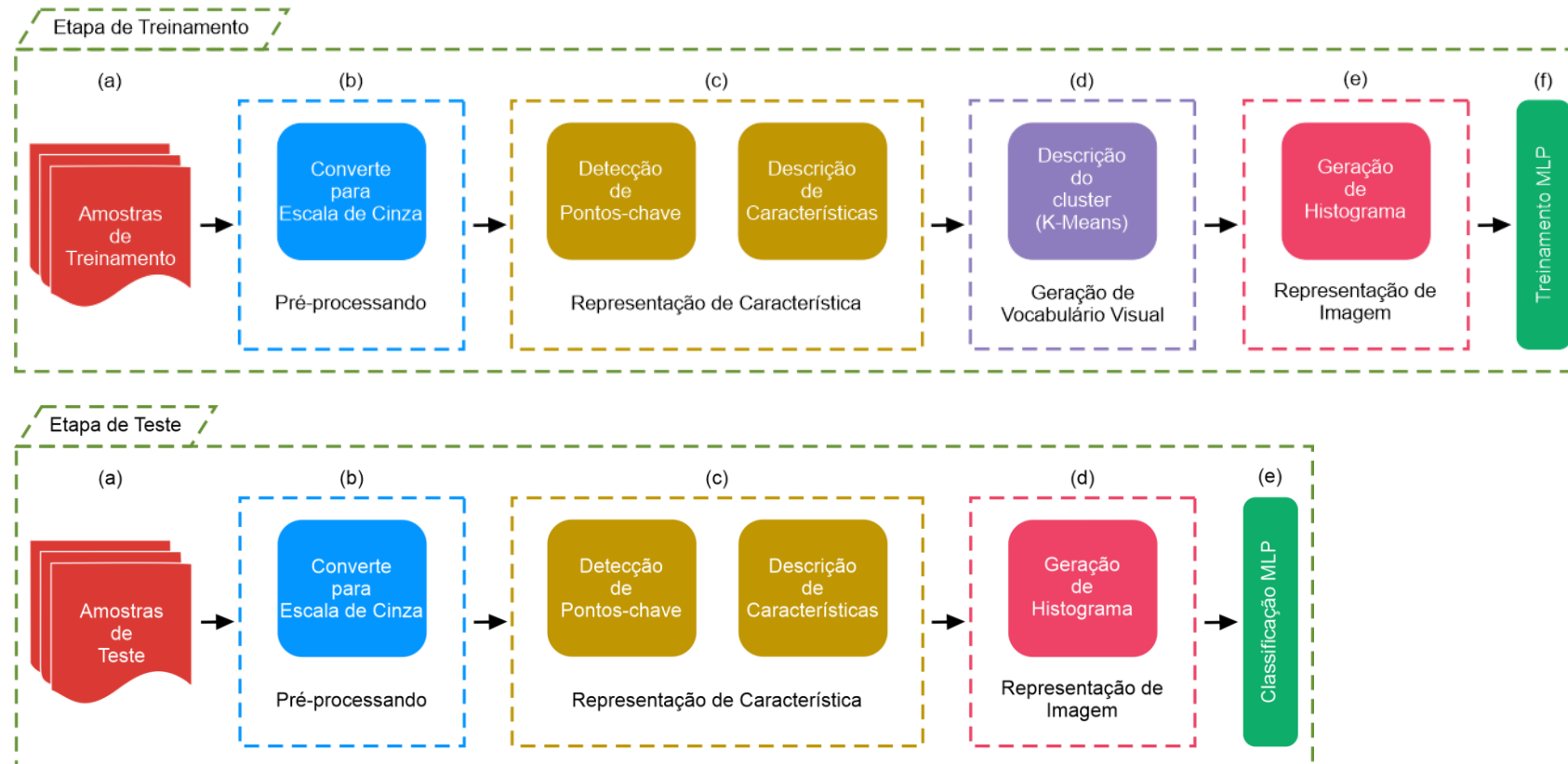
Detecção e Correspondência de Características



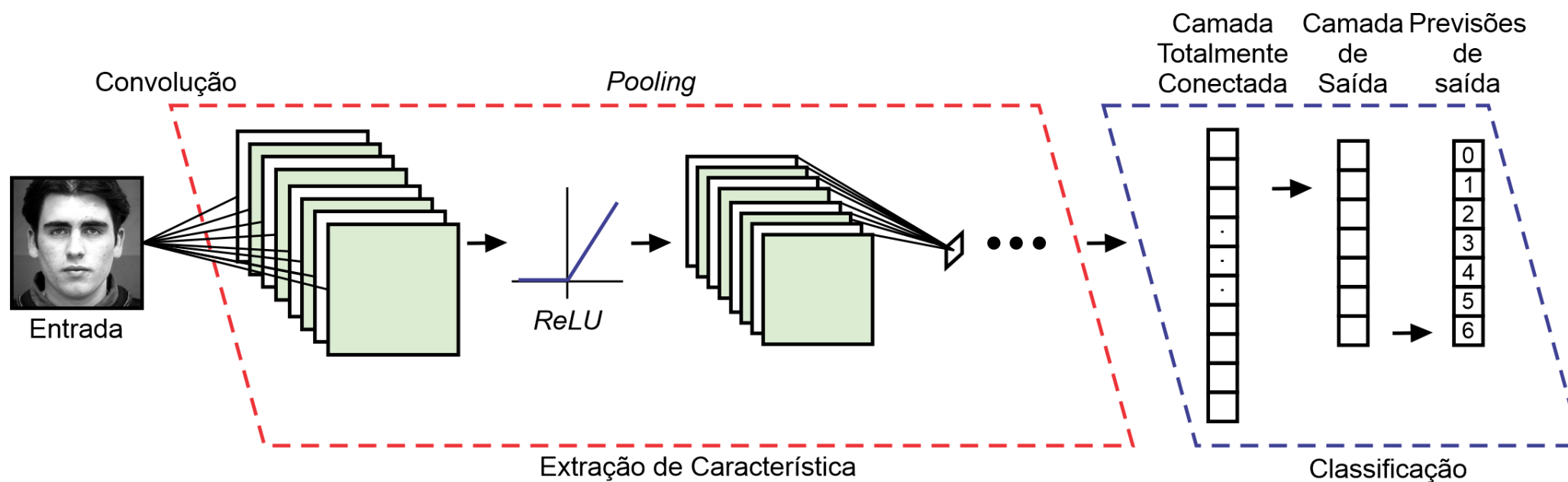
Detecção e Correspondência de Características



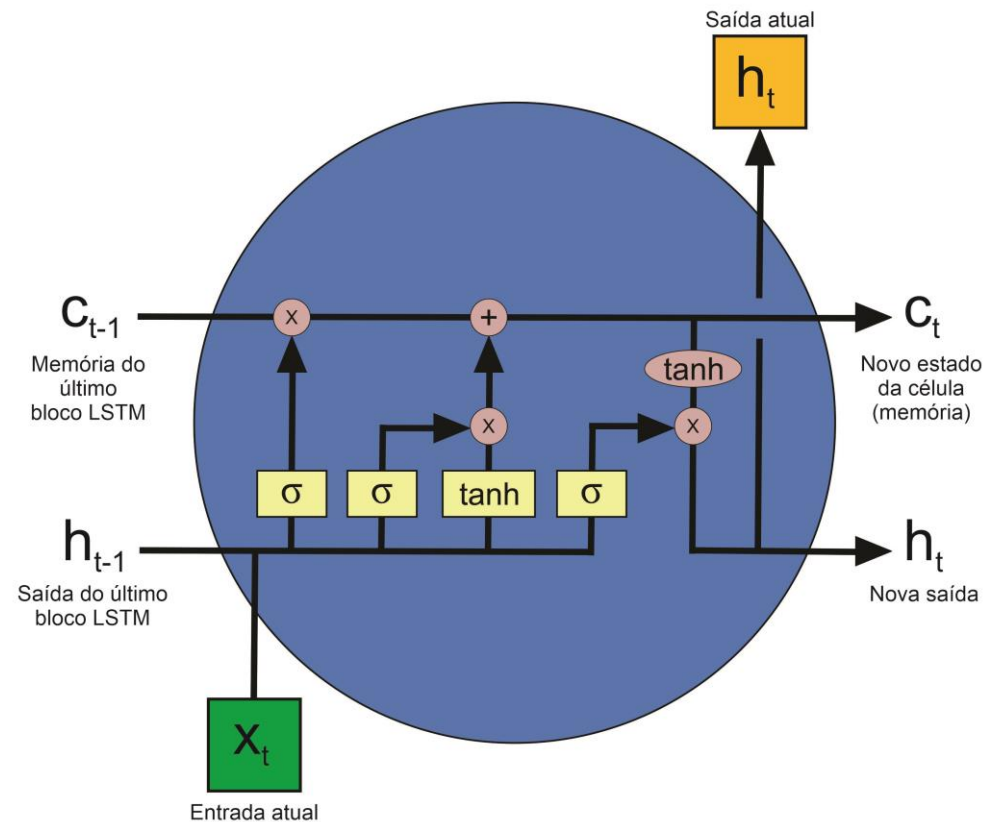
Saco de Características Visuais



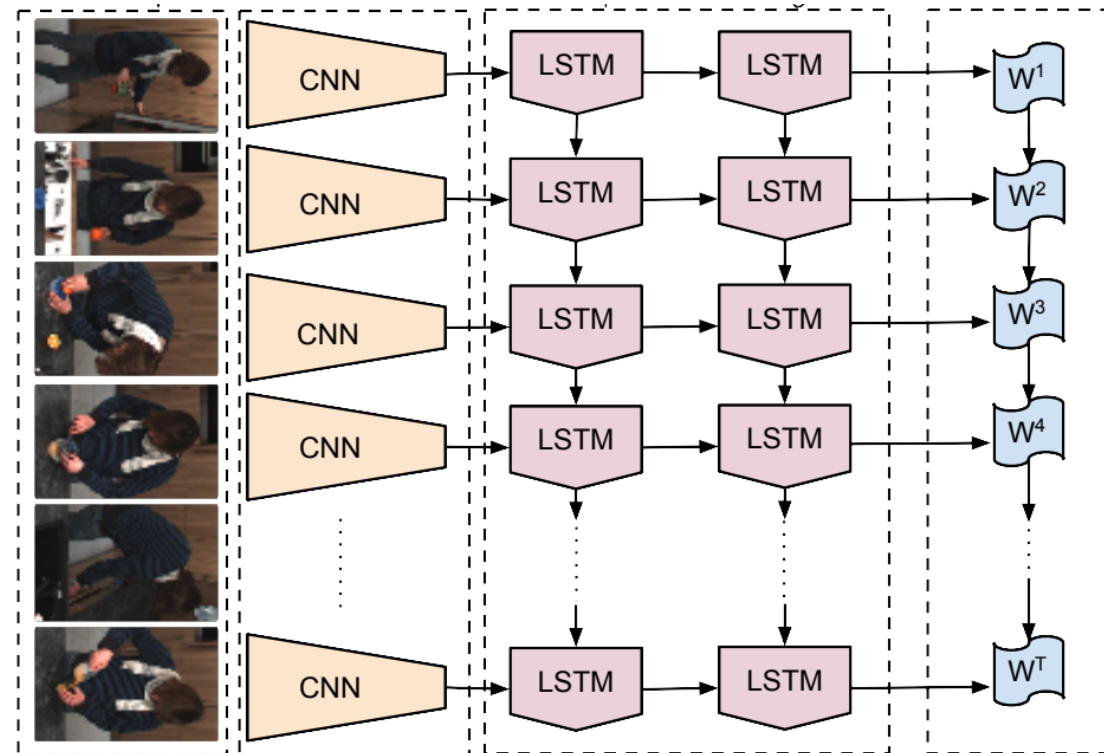
Redes Neurais Convolucionais



Redes Neurais Recorrentes



Modelo Híbrido de Arquitetura de Rede Neural Artificial



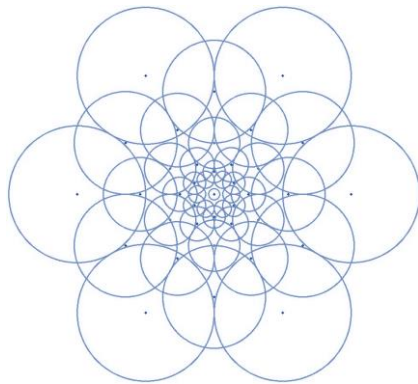
Trabalhos Relacionados

- A eficiência entre descritores pode ser vistas em diversos trabalhos realizado pela comunidade científica, onde foram aplicadas diversas análises e avaliações [1], [2], [3], [4].
- Em Dai et al. [5] os autores fornecem um estudo comparativo de três classes distintas de Descritores de Recursos Locais, são eles: recursos artesanais, Redes Neurais Convolucionais treinadas e Redes Neurais Convolucionais pré-treinadas para avaliar sua eficiência para a tarefa de correspondência de pontos-chave em aplicações de Robótica, levando em consideração a capacidade dos descritores em lidarem com alterações condicionais.

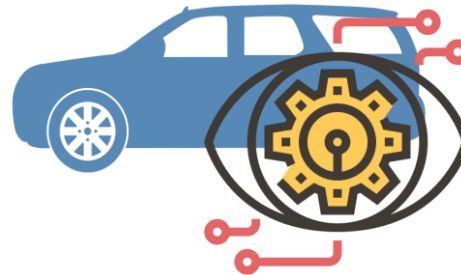
Trabalhos Relacionados

- Loo et al. [6] aprimora o mapeamento da Odometria Visual através de uma Rede Neural Convolutacional de previsão de profundidade de imagem única.
- Zhang et al. [7] apresenta uma nova abordagem para a solução do problema de detecção de fechamento de loop para sistemas VSLAM baseado em Redes Neurais Convolucionais

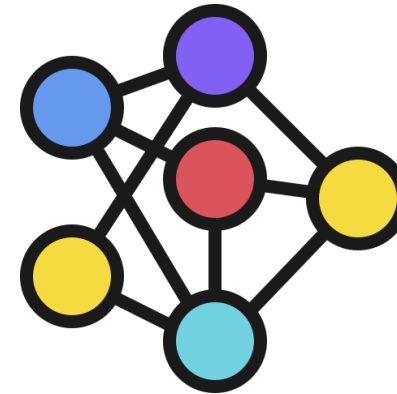
Destaque



Descritores de Recursos
Locais e Descritores
Binários Locais



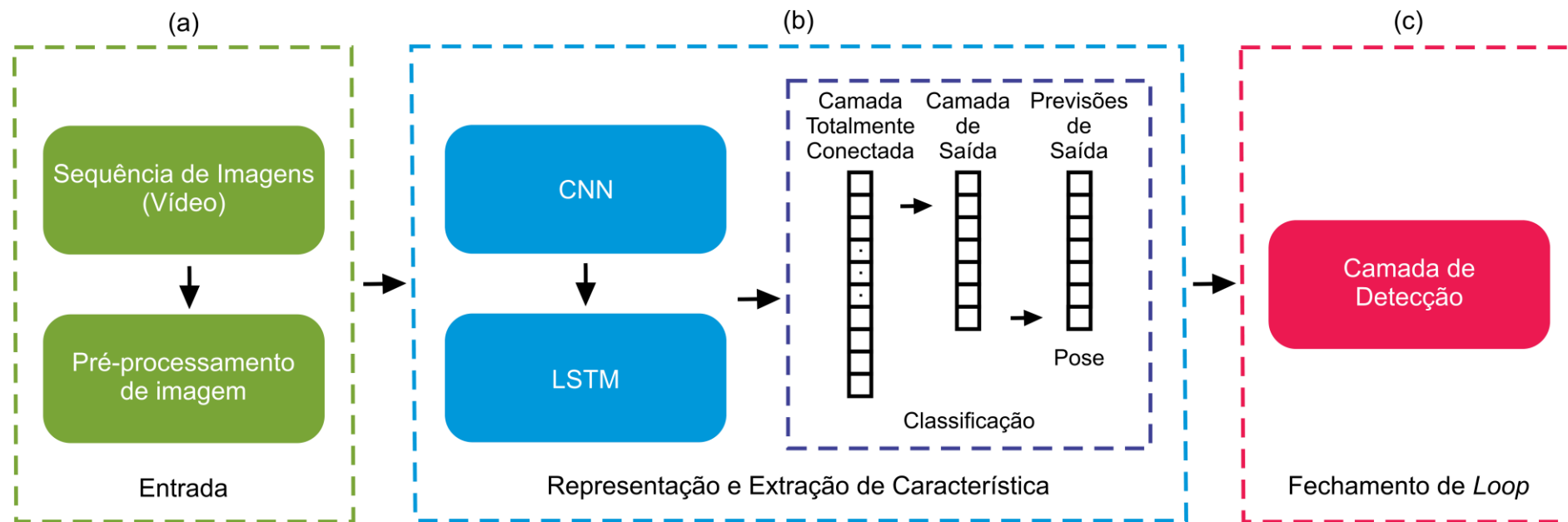
SLAM baseado
em visão



Redes Neurais
Convolucionais + Detecção
de Fechamento de Loop

A Proposta

A Proposta



Viabilidade

Linguagem de Programação e Bibliotecas



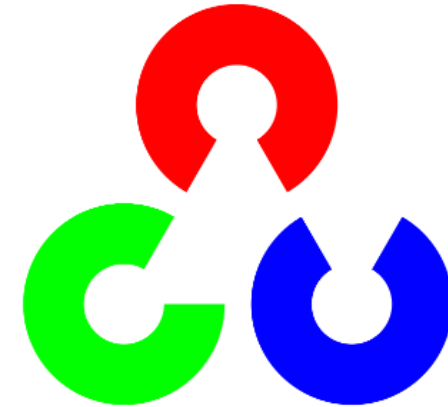
Python



TensorFlow



Keras



OpenCV

Kit de Desenvolvimento NVIDIA Jetson Nano



Experimentos e Resultados

Avaliação comparativa entre Descritores

- Avaliação de uma abordagem de Saco de Características Visuais extraíndo características através de Descritores de Recursos Locais e Descritores Binários Locais no microcomputador Jetson Nano para as tarefas de reconhecimento e classificação em seis conjuntos de dados visuais por meio do classificador Multilayer Perceptron:
 - MNIST, JAFFE, Extended CK+, FEI, CIFAR-10, e FER-2013.

Avaliação comparativa entre Descritores

- Descritores de Recursos Locais:
 - SIFT, SURF, and KAZE.
- Descritores Binários Locais:
 - BRIEF, ORB, BRISK, AKAZE, e FREAK.

Etapa de Treinamento

Etapa de Representação de Características					
Algoritmos	<i>Dataset</i>	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

Geração de Vocabulário Visual					
Algoritmos	<i>Dataset</i>	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

Etapa de Representação de Imagem					
Algoritmos	<i>Dataset</i>	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

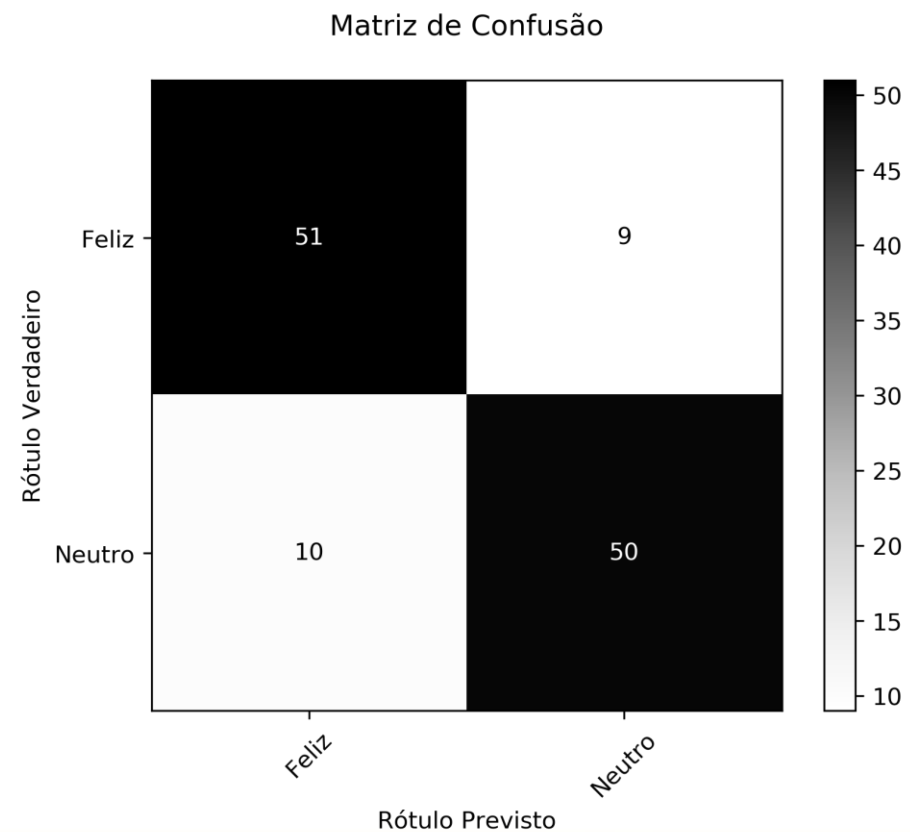
Etapa de Teste

Etapa de Representação de Características e Representação de Imagem					
Algoritmos	<i>Dataset</i>	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

Algoritmos	<i>Dataset</i>	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



Rede Neural Convolucional de Descritores

- Foi investigado técnicas e métodos que pudessem ser utilizadas para o desenvolvimento da reformulação de filtros convolucionais de uma arquitetura de Rede Neural Convolucional tradicional através de um Descritor Binário Local, e consequentemente, utilizadas também na implementação da adaptação de Rede Neural Convolucional, apresentando o novo design de camadas para a arquitetura proposta.

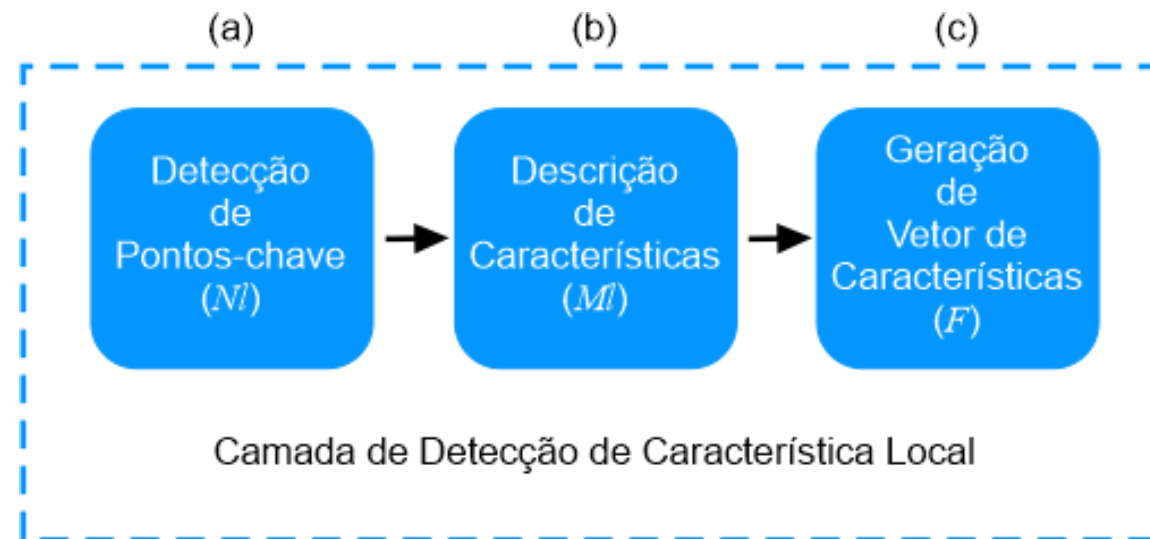
Rede Neural Convolucional de Descritores

- Assim, propomos a reformulação das camadas convolucionais por meio de Descritores Binários Locais.
- Com isto, apresentamos duas novas camadas para arquiteturas profundas: (a) Detecção de Característica Local e (b) Convolução de Descritor Local, como uma alternativa viável e eficiente às camadas convolucionais.

Detecção de Característica Local

- Primeira camada da arquitetura proposta.
- Responsável por executar a (a) detecção de pontos-chave Nl no conjunto de dados de treinamento e, em seguida; (b) os descritores são calculados, portanto, para cada imagem, N pontos-chave serão detectados e para cada ponto-chave, M descritores serão calculados, com isso; (c) é gerando um vetor de características F .

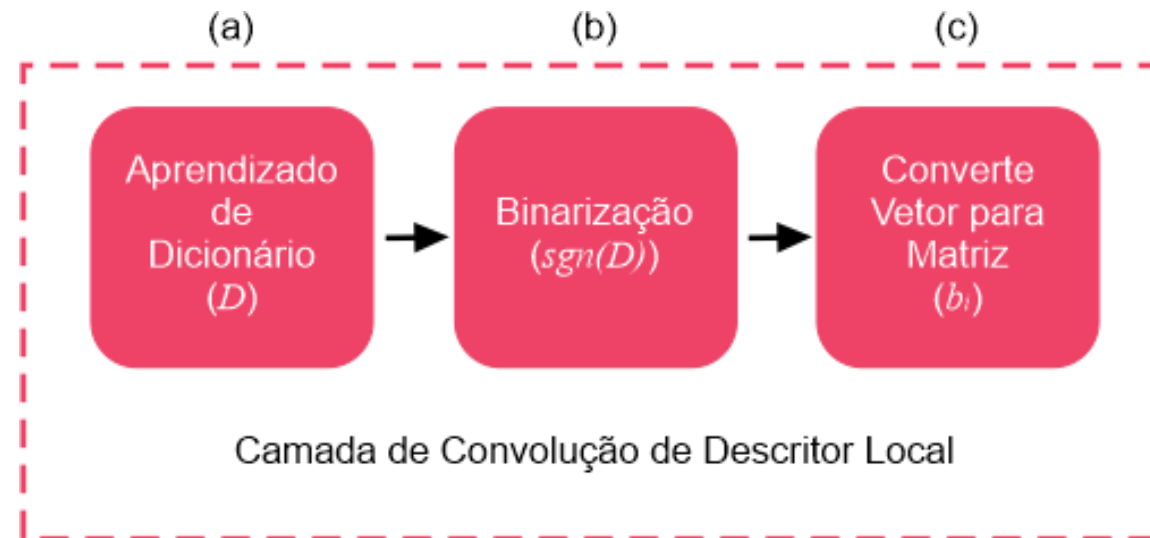
Detecção de Característica Local



Convolução de Descritor Local

- Alternativa à camada convolucional de uma Rede Neural Convolucional tradicional.
- Consiste em (a) reduzir a dimensionalidade por meio do algoritmo Aprendizado de Dicionário aplicado ao vetor de característica gerado pela camada de Detecção de Característica Local anterior, gerando um dicionário esparsos D (conjunto de átomos), ou seja, gerando um vetor de características esparsos a partir do descritor; (b) seguida por um processo de binarização, neste caso, a função Sinal, transformando a natureza do dicionário esparsos D , em binário e esparsos $\text{sgn}(D)$; (c) seguida ainda, por uma transformação do dicionário esparsos, que nada mais é do que a transformação de um vetor de características em uma matriz quadrada, e, finalmente, é definido para o centro de todas as matrizes quadradas, um valor negativo (-1).

Convolução de Descritor Local



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

Comportamento da operação de convolução

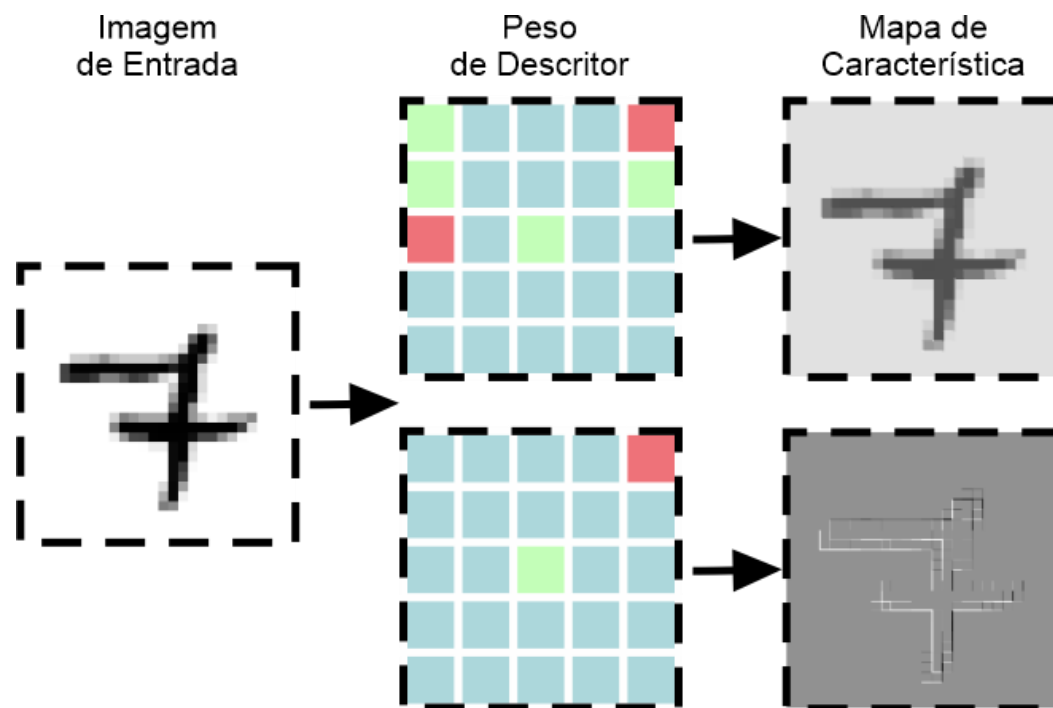
0	0	0	0	0	0	0
0	60	113	56	139	85	0
0	73	121	54	84	128	0
0	131	99	70	129	127	0
0	80	57	115	69	134	0
0	104	126	123	95	130	0
0	0	0	0	0	0	0

Kernel

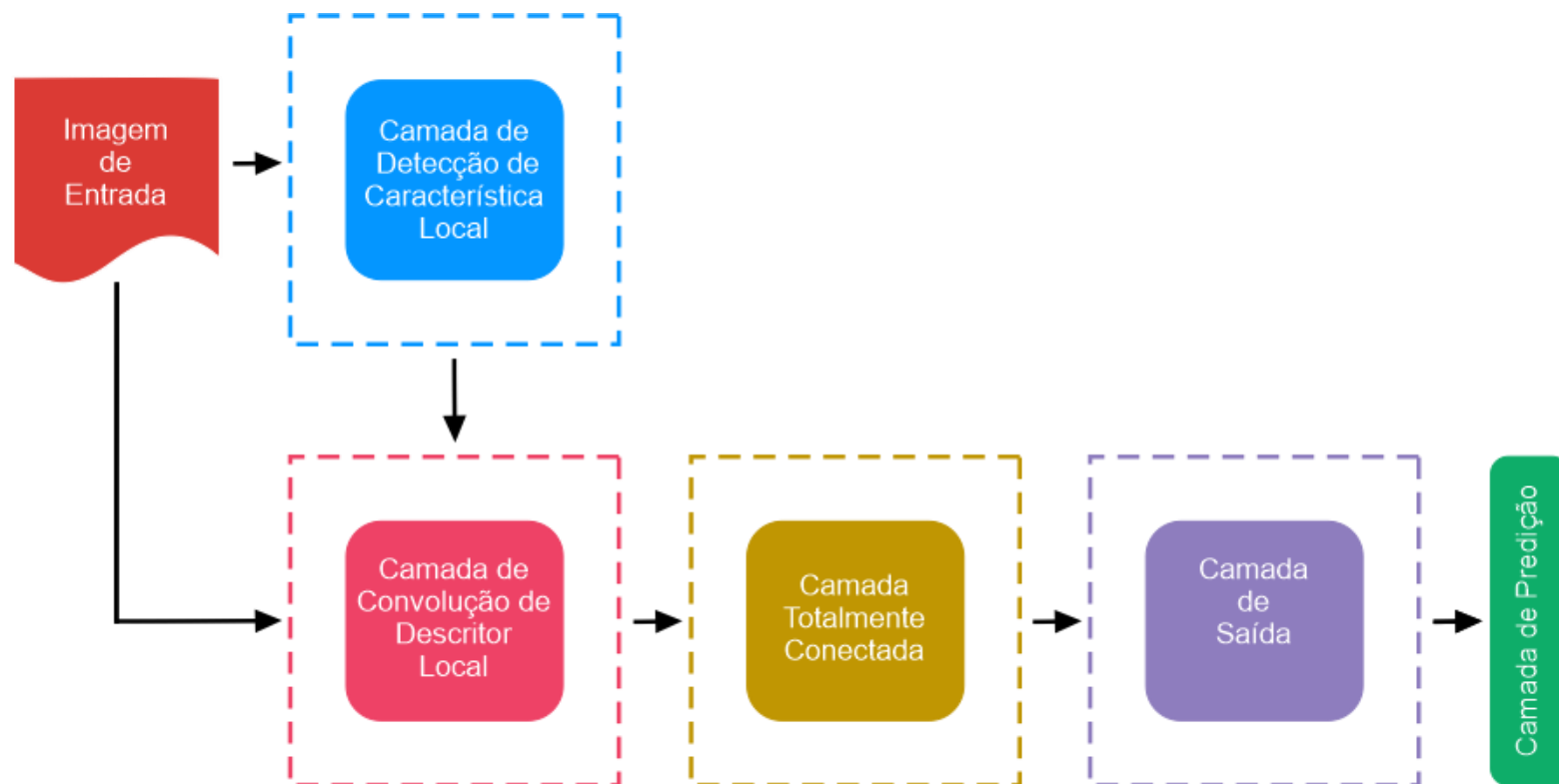
0	-1	0
-1	5	-1
0	-1	0

114				

Comportamento da operação de convolução



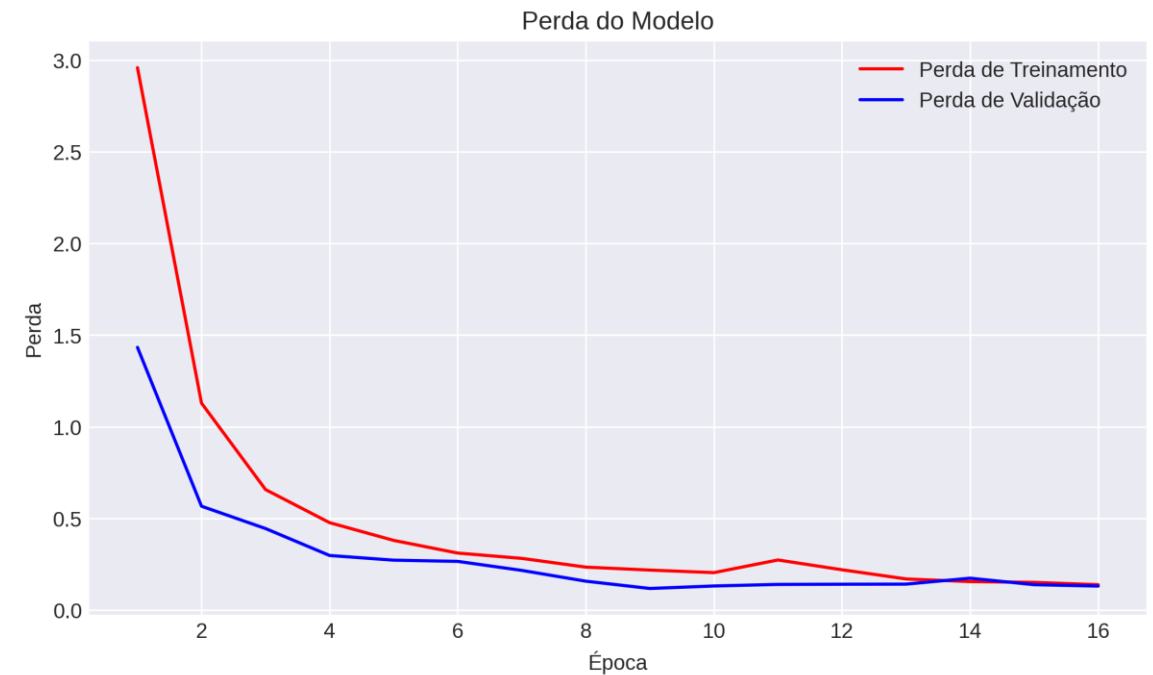
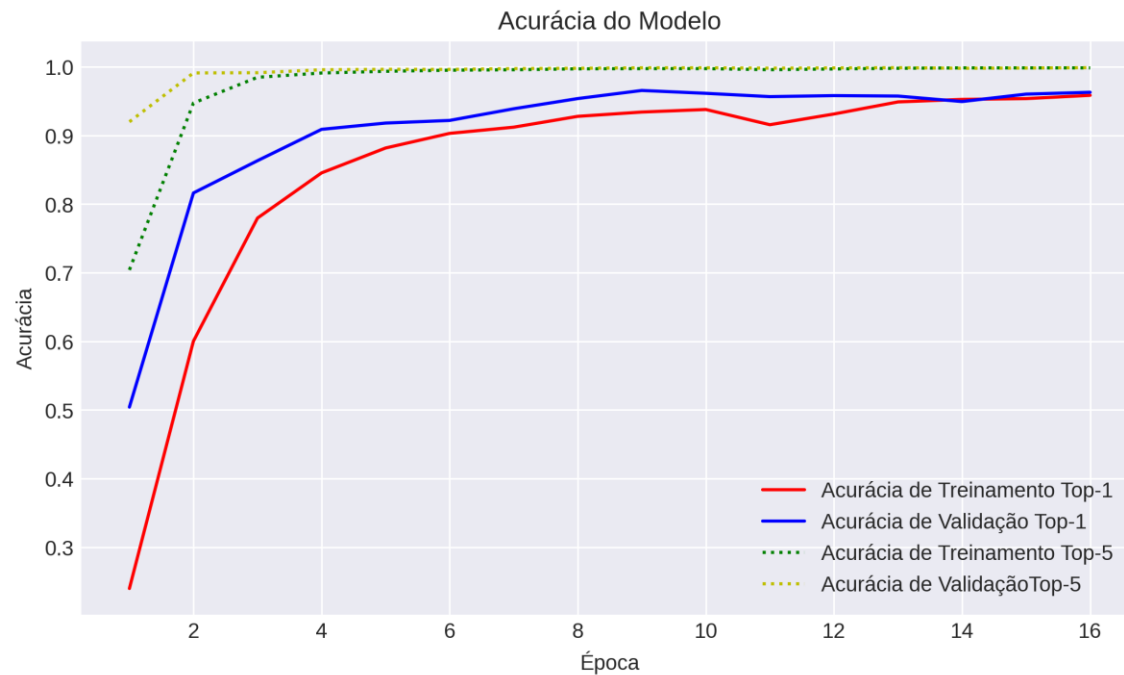
Arquitetura básica de Rede Neural Convolucional de Descritores



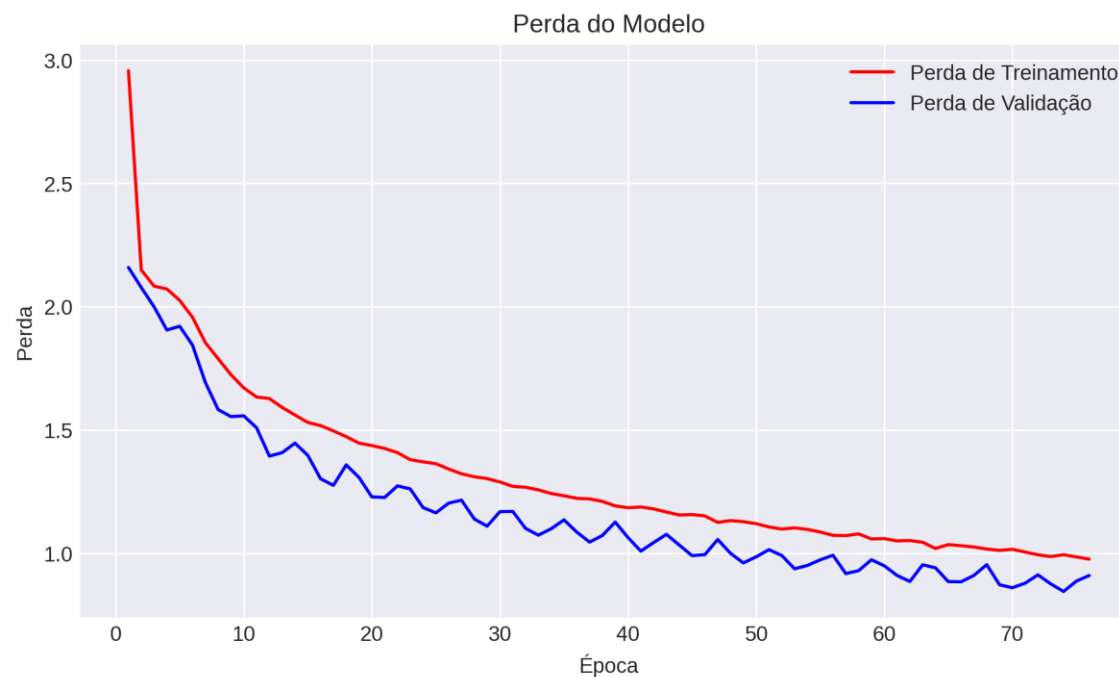
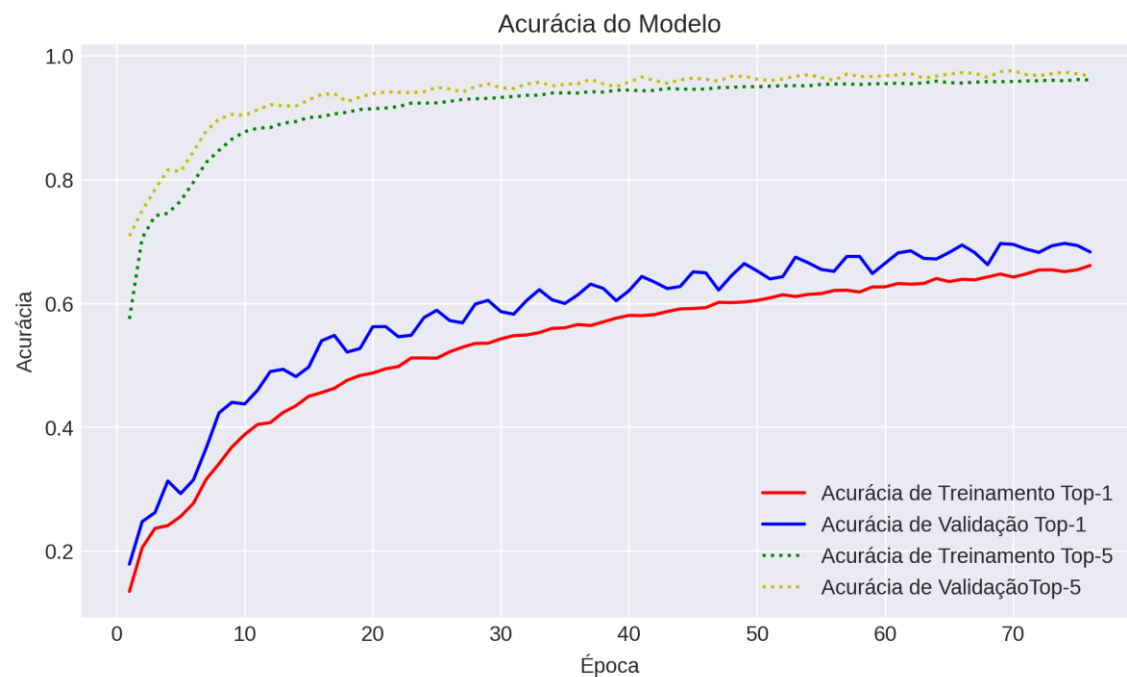
Rede Neural Convolucional de Descritores

- Conceitualmente, a DescNet proposta neste trabalho pode ser facilmente implementada em qualquer estrutura de aprendizado profundo existente.
- Como topologia da rede, para estes experimentos, foi utilizado a arquitetura ResNet-152.
- Portanto, a mesma arquitetura utilizada para implementar a Rede Neural Convolucional tradicional serviu como linha de base para a implementação da DescNet.

Resultados no conjunto de dados visuais MNIST



Resultados no conjunto de dados visuais CIFAR-10



Comparação com Linha de Base

- A Tabela consolida os resultados obtidos da DescNet de melhor desempenho em comparação com sua linha de base (ResNet-152) correspondente.

Conjunto de dados visuais	DescNet				Linha de Base			
	Top-1	Top-25	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

Conclusão

- Experimentos foram realizados para comparar sua eficiência com um modelo de Rede Neural Convolutacional tradicional.
- Alcançamos resultados razoáveis sobre a camada convolutacional padrão em conjuntos de dados visuais competitivos (MNIST e CIFAR-10) enquanto permite economia no número de parâmetros do modelo e, conseqüentemente, 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.

Referências

Referências

- [1] PATEL, Akash et al. Performance analysis of various feature detector and descriptor for real-time video based face tracking. International Journal of Computer Applications, v. 93, n. 1, 2014.
- [2] CHATOUX, Hermine; LECELLIER, François; FERNANDEZ-MALOIGNE, Christine. Comparative study of descriptors with dense key points. In: 2016 23rd International Conference on Pattern Recognition (ICPR). IEEE, 2016. p. 1988-1993.
- [3] BAYRAKTAR, Ertuğrul; BOYRAZ, Pinar. Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics. Turkish Journal of Electrical Engineering and Computer Science, v. 25, n. 3, p. 2444-2454, 2017.
- [4] TAREEN, Shaharyar Ahmed Khan; SALEEM, Zahra. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In: 2018 International conference on computing, mathematics and engineering technologies (iCoMET). IEEE, 2018. p. 1-10.

Referências

- **[5]** Dai, Z., Huang, X., Chen, W., He, L., & Zhang, H. (2019, May). A Comparison of CNN-Based and Hand-Crafted Keypoint Descriptors. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 2399-2404). IEEE.
- **[6]** Loo, S. Y., Amiri, A. J., Mashohor, S., Tang, S. H., & Zhang, H. (2019, May). CNN-SVO: Improving the mapping in semi-direct visual odometry using single-image depth prediction. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 5218-5223). IEEE.
- **[7]** Zhang, X., Su, Y., & Zhu, X. (2017, September). 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.

DESAFIOS NA CONSTRUÇÃO DE UM SISTEMA DE NAVEGAÇÃO DE ROBÔS AUTÔNOMOS MÓVEIS BASEADO EM VISÃO: DETECÇÃO DE FECHAMENTO DE LOOP

Alexandra Miguel Raibolt da Silva
(raibolt@ime.eb.br)

OBRIGADA!

