CIÊNCIA DA COMPUTAÇÃO

Geração de Imagens a Partir de Sentença Descritiva

Aluno: Augusto Turolla, Mateus Davi Silva, Igor do Nascimento Alves e Alexandre Kenjy de Siqueira Kumagai
(igor.nascimento.flipe@gmail.com, gutoturolla@hotmail.com, alexandrekenjy@gmail.com, mateus_rmn@hotmail.com)

Orientador: Prof. Paulo S. Rodrigues (psergio@fei.edu.br)

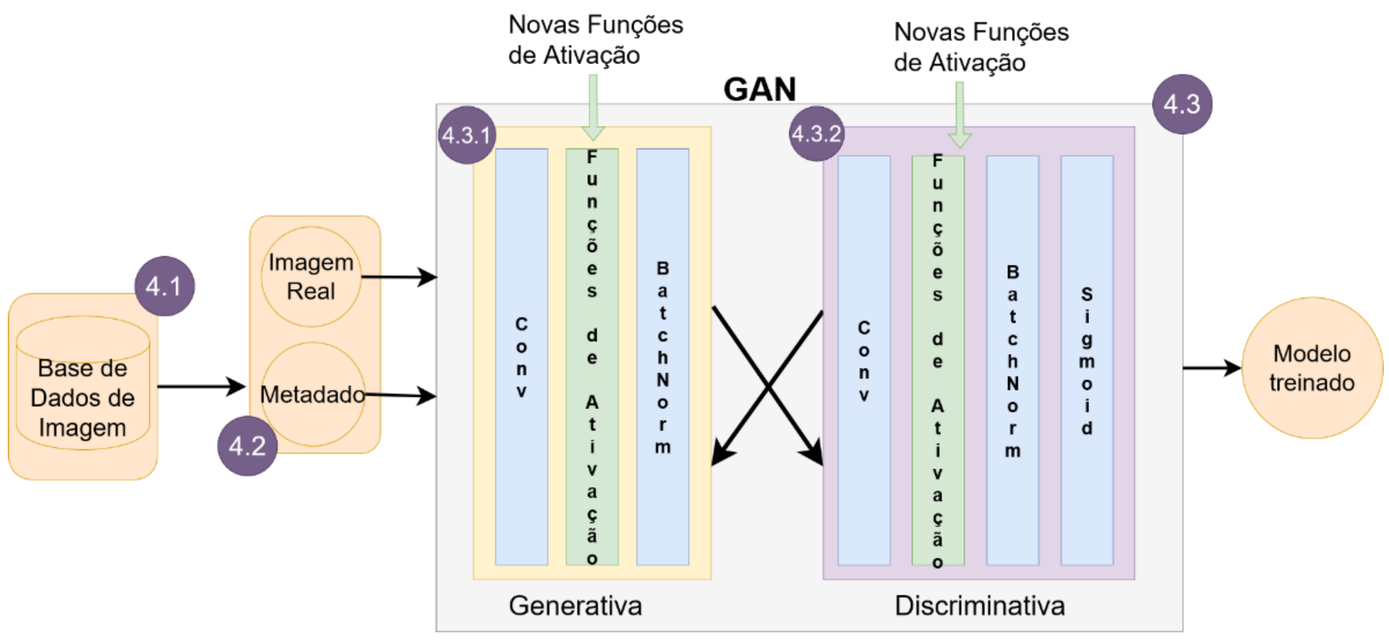
RESUMO

Com o avanço tecnológico, sobretudo nas áreas de Inteligência Artificial(IA), Processamento de Linguagem Natural(PLN) e Aprendizado Profundo existem expectativas promissoras para futuras tarefas que a computação possivelmente será capaz de alcançar. Um destes problemas é o da capacidade da computação de ilustrar contextos diversos. Alguns exemplos deste problema estão relacionados à indústria da literatura, cinematografia e de jogos ou de criar cenas para áreas comercial e publicitária. Neste trabalho, é proposto um modelo capaz de gerar imagens a partir de sentenças textuais. Para construir o modelo é utilizada uma rede neural para gerar imagens, chamada de rede generativa adversarial (GAN) O projeto obteve sucesso em gerar imagens que se assemelham ao contexto urbano apresentando uma assertividade de 27% através da função ReLU como ativadora da rede neural.

METODOLOGA

A metodologia proposta neste trabalho é composta de quatro etapas. A primeira etapa consiste em pré-processamento do dataset utilizado. Na segunda etapa ocorre o tratamento das sentenças de entrada através da técnica de sumarização de texto, utilizando o sistema Attention-based Summarization. Nessa segunda etapa, são extraídos atributos de textos para serem processados pelas redes GAN-CLS. Na etapa seguinte, através de decodificação, os atributos de texto possuirão um formato que possibilitará o reconhecimento através da arquitetura GAN-CLS. Por fim, após a imagem gerada for aceita, serão realizadas verificações através de validações humanas, técnica comumente utilizada em avaliações de redes GANs, devido à falta de métricas específicas para estes modelos.

Diagrama esquemático do treinamento da metodologia proposta.



RESULTADOS

A figura ao lado mostra as imagens geradas pelos 5 modelos cada um com sua respectiva função de ativação a partir de uma mesma descrição. Nela, é possível notar que as imagens são bem diferentes uma das outras e que alguns modelos se saíram melhor em representar detalhes e outros em criar silhuetas da cidade. As imagens geradas pelos modelos que utilizaram as funções ReLU e Softsign tiveram destaque na avaliação de comparação feita.

Tabela 4 – Tabela comparativa das imagens geradas sobre o contexto urbano

Dataset - MS COCO - Urbano	
Texto: A busy city street with lots of cars	
Função de Ativação	Imagem
ReLU	
ReLU6	
PReLU	
Sigmóide	
Soft Sign	

CONCLUSÃO

O projeto proposto obteve 27% e 24% de assertividade. O maior desafio no momento ainda é a dificuldade na geração de cenas com múltiplos objetos, mesmo não citados no texto mas que fazem parte do contexto. Sendo assim, a produção de trabalhos futuros é uma via de diversas possibilidades, com oportunidades de pesquisas em diversas áreas.