



# Aprendizado Profundamente Supervisionado em Grafos Aplicado a Pecuária de Precisão

Programa de Pós-Graduação Stricto Sensu - Doutorado em Computação

INOVISAO

Universidade Federal de Mato Grosso do Sul

Janeiro de 2019

Campo Grande - MS

---

# 1 Antecedentes e Justificativa

Atualmente, as redes neurais convolucionais têm obtido uma grande atenção devido aos seus resultados em bancos de imagens como a ImageNet [Russakovsky et al., 2015]. As redes neurais convolucionais consistem em camadas convolucionais, que utilizam o algoritmo de *back-propagation* para aprender os parâmetros/filtros de cada camada. Estes filtros permitem que o modelo aprenda os padrões visuais locais importantes, sendo ainda invariante à translação. Apesar destas grandes contribuições, existe alguma limitações em relação as redes neurais convolucionais:

1. Falta de transparência e discriminação das características aprendidas nas camadas convolucionais;
2. A estrutura dos filtros precisam ser regulares (e.g., matrizes quadradas  $3 \times 3$  ou  $5 \times 5$ ).

Para tentar solucionar a primeira limitação, as linhas de pesquisa que tem apresentado resultados promissores são as chamadas Redes Profundamente Supervisionadas (Deeply-supervised networks - DSN) [Lee et al., 2015, Wang et al., 2015, Mostafa et al., 2018, Bazi e Melgani, 2018]. Esses trabalhos surgiram com o intuito de treinar as camadas escondidas de forma supervisionada. Assim, as camadas escondidas também recebem supervisão e não somente a última camada, o que aumenta o controle do aprendizado das camadas internas. Com o aprendizado profundamente supervisionado, [Bazi e Melgani, 2018] também mostraram uma menor tendência de um sobreajuste (*overfitting*) e a necessidade de poucas imagens de treinamento quando comparado com o método tradicional.

Para tentar solucionar a segunda limitação, uma nova linha de pesquisa têm surgido: aprendizagem profunda geométrica [Bronstein et al., 2017]. Esta metodologia tem como o objetivo

---

generalizar a aprendizagem profunda para amostras regulares e não regulares. Uma das principais amostras não regulares são os grafos, portanto estudos de como utilizar grafos como estrutura das amostras no treinamento das CNNs têm obtido bastante atenção. Este problema não é trivial e possui diversos desafios como: domínio irregular, estruturas e tarefas variadas e, por fim, escalabilidade e paralelização.

A partir destes direcionamentos, este projeto tem como objetivo propor métodos de redes neurais profundamente supervisionadas. A ideia é que os filtros das camadas convolucionais sejam substituídos por classificadores SVMs, conforme proposto por [Bazi e Melgani, 2018]. Entretanto, no trabalho de [Bazi e Melgani, 2018], os SVMs foram treinados de forma independente (sem o uso do *backpropagation*) e somente para duas classes. Por outro lado, pretende-se estudar uma forma de treinar estes classificadores de forma dependente, como acontece no *backpropagation*. Também deseja-se propor a ideia de utilizar SVMs capazes de classificar múltiplas classes, o que não foi avaliado no trabalho original.

Em geral, as entradas das redes neurais convolucionais tradicionais e profundamente supervisionadas são estruturas regulares como grades e imagens. Porém, existem diversos extratores de características na literatura (e.g., superpixel [Achanta et al., 2012]) com bom desempenho cuja as suas saídas são grafos. Conforme descrito na literatura, os grafos melhoram a representação pois inclui relações entre os nós que vão além da relação espacial da imagem. Dessa forma é importante que as redes neurais convolucionais sejam capazes de convoluir através de estruturas não regulares. Portanto, este projeto também propõe o uso de estruturas não regulares na convolução de uma rede.

Os métodos proposto serão aplicados em problemas da pecuária de precisão. A pecuária é um dos segmentos que mais se destaca na economia brasileira. O Brasil detém o maior rebanho bovino comercial do mundo, com mais de 226 milhões de cabeças, e o abate supera 35 milhões de animais por ano. Em 2016, o PIB do agronegócio correspondeu a 24% do PIB do Brasil, sendo que a pecuária representou 31% do agronegócio. O Brasil produziu aproximadamente 10 milhões de toneladas de carne bovina e exportou mais de 1,630 milhão de toneladas para

---

diversos países em 2017 [Kist, 2017]. Apesar dos números elevados, o sistema nacional ainda apresenta índices produtivos pouco representativos devido à baixa produtividade média por área comparada a outros países. Para melhorar estes índices é necessário adotar tecnologias por meio da pecuária de precisão que permitam incrementos em produtividade e maior rentabilidade ao produtor. A pecuária de precisão consiste na medição de diferentes parâmetros dos animais e do ambiente, a modelagem desses dados e o seu uso para tomada de decisões. Para tornar as medições automáticas, métodos computacionais com base em imagens estão sendo cada vez mais estudados e propostos na literatura [Frost et al., 1997]. Estes métodos utilizam a visão computacional cujo objetivo é construir sistemas artificiais para interpretar imagens. Dois problemas da pecuária de precisão serão abordados neste projeto: estimar o peso do animal e tipificação de carcaça através de imagens. Ambos problemas são considerados desafiadores devido ao ambiente não controlado.

Por meio deste projeto, ambas as áreas, visão computacional e agropecuária, serão beneficiadas. Por um lado, novos desafios são apresentados aos métodos de visão computacional, tornando possível o aperfeiçoamento e a criação de novos métodos para suprir as necessidades e aumentar a taxa de acerto. Por outro lado, os algoritmos e métodos desenvolvidos podem contribuir para novas tecnologias na agropecuária de precisão, aumentando a produtividade e reduzindo os custos da produção.

## 2 Objetivos

### 2.1 Objetivo Geral

O objetivo geral deste projeto é propor uma rede neural convolucional profundamente supervisionada com classificadores SVM como filtros das camadas convolucionais. Além disso, pretende-se propor a utilização de formas não geométricas (e.g. grafos e superpixels) como entrada das camadas convolucionais, diferente das CNNs tradicionais que usam formas geométricas, como grades e imagens. Os métodos propostos serão aplicados em problemas da pecuária de

precisão tais como estimativa do peso de bovinos e classificação da qualidade da carcaça de bovinos.

## 2.2 Objetivos Específicos

1. Criação dos bancos de imagens dos problemas de pecuária de precisão;
2. Investigar e propor formas não geométricas para entrada das camadas convolucionais;
3. Desenvolvimento e implementação de métodos profundamente supervisionados;
4. Aplicação dos métodos propostos em problemas de pecuária de precisão;
5. Divulgação dos resultados obtidos por meio de artigos científicos.

## 3 Revisão de Literatura

Esta seção apresenta os trabalhos da literatura relacionados com redes neurais convolucionais profundamente supervisionadas e e aprendizado profundo em grafos.

### 3.1 Redes Neurais Convolucionais

As redes neurais convolucionais consistem em camadas convolucionais, que utilizam o algoritmo de *backpropagation* para aprender os parâmetros/filtros de cada camada. Desde a sua criação, a CNN tem sido caracterizada por três propriedades básicas, as conexões locais, o compartilhamento de peso e o *pooling* local. As duas primeiras propriedades permitem que o modelo aprenda os padrões visuais locais importantes com menos parâmetros ajustáveis que um modelo totalmente conectado, e a terceira propriedade prepara a rede para possuir invariância à translação [LeCun et al., 1989].

Uma das primeiras propostas de redes neurais convolucionais é a rede LeNet-5 descrita por LeCun et al. [LeCun et al., 1989] para o reconhecimento óptico de caracteres. Comparada às

redes convolucionais profundas atuais, a rede foi relativamente modesta devido aos recursos computacionais limitados da época e aos desafios do treinamento de algoritmos para redes com mais camadas. Embora houvesse muito potencial em redes convolucionais mais profundas, só recentemente elas se tornaram predominantes, devido o aumento do poder computacional atual, da quantidade de dados para treinamento disponíveis na internet e do desenvolvimento de métodos mais eficazes para a formação de tais modelos.

Um exemplo recente e notável do uso de redes convolucionais profundas para classificação de imagens é o desafio Imagenet [Krizhevsky et al., 2012a] em que uma CNN obteve um erro consideravelmente menor comparado com o erro das abordagens tradicionais de visão computacional. Redes neurais convolucionais também obtiveram recentemente sucesso para diferentes aplicações, incluindo estimação de pose humana [Toshev e Szegedy, 2014], análise de faces [Luo et al., 2012], detecção de pontos-chave facial [Sun et al., 2013], reconhecimento de voz [Graves et al., 2013] e classificação de ação [Karpathy et al., 2014]. Apesar deste grande sucesso, alguns problemas sobre o aprendizado profundo surgiram, tais como:

- Falta de transparência e discriminação das características nas camadas escondidas da rede [Zeiler e Fergus, 2014].
- Pouco conhecimento matemático sobre o comportamento do algoritmo [Eigen et al., 2013].
- Necessidade de grandes bancos de imagens no treinamento [Hinton et al., 2012].
- Complexidade para ajustar os parâmetros da rede [Krizhevsky et al., 2012b].

Para tentar resolver estes problemas, técnicas têm surgido como a utilização de uma rede neural convolucional pré-treinada [Dahl et al., 2012] e o aumento de dados [Ciresan et al., 2012]. Porém, apesar de avanços nos estudos sobre redes neurais convolucionais, ainda existe uma grande dificuldade em entender o seu processo de treinamento. Uma das linhas de pesquisa que tem apresentado resultados promissores são as chamadas Redes Profundamente Supervisionadas (Deeply-supervised networks - DSN) [Lee et al., 2015, Wang et al., 2015, Mostafa et al., 2018,

Bazi e Melgani, 2018]. Esses trabalhos surgiram com o intuito de treinar as camadas escondidas de forma supervisionada. Assim, as camadas escondidas também recebem supervisão e não somente a última camada, o que aumenta o controle do aprendizado das camadas internas. Com o aprendizado profundamente supervisionado, [Bazi e Melgani, 2018] também mostraram uma menor tendência de um sobreajuste (*overfitting*) e a necessidade de poucas imagens de treinamento quando comparado com o método tradicional.

Um dos primeiros trabalhos nessa linha de pesquisa foi proposto por [Lee et al., 2015]. Os autores propõem uma rede profundamente supervisionada que tenta minimizar o erro de classificação enquanto torna o processo de aprendizado das camadas escondidas direto e transparente. Para isso, um classificador SVM é adicionado em cada camada escondida e o seu erro é propagado junto como o *backpropagation*, otimizando a função de perda. Duas arquiteturas diferentes foram testadas mudando a última camada, onde uma utiliza o *softmax* e a outra o SVM.

Apesar dos bons resultados, adicionar um classificador para cada camada escondida se torna inviável para redes com muitas camadas. Dessa forma, [Wang et al., 2015] introduziram a ideia de incluir uma regra para adicionar ou não este classificador em uma determinada camada escondida. Inicialmente, a rede é treinada com supervisão apenas na última camada e os gradientes das camadas escondidas são monitorados. As camadas cujo o gradiente médio torna-se pequeno (e.g., menor que  $10^{-7}$  de acordo com os autores) recebem uma supervisão. Em uma CNN com oito camadas, foi necessário incluir uma supervisão após a quarta camada. A supervisão consiste em pequenas redes compostas por uma camada convolucional seguida por várias camadas totalmente conectadas e o *softmax*.

A grande maioria dos trabalhos atuais usam o *backpropagation* para o treinamento das CNNs, que requer que as primeiras camadas aguardem a propagação do erro das camadas mais profundas. Entretanto, [Mostafa et al., 2018] mostraram que esta sincronia dos erros não ocorre em redes neurais biológicas. Sendo assim, eles propuseram adicionar uma rede com múltiplas camadas para prever os erros locais das camadas escondidas e assim tornar o treinamento das camadas independentes. Os erros são gerados diretamente das classes de treinamento usando

um classificador para cada camada. Assim ao invés de minimizar uma função de perda global, o treinamento minimiza várias funções de perda locais.

Por outro lado, [Bazi e Melgani, 2018] utilizaram classificadores SVMs para substituir os filtros das camadas convolucionais. Para isso, cada camada convolucional é iniciada com um conjunto de SVMs lineares como filtros e treinados de forma independente, sem a utilização do *backpropagation*. Para o treinamento dos classificadores, amostras (*patches*) do objeto de interesse são introduzidos de forma supervisionada. A ideia de que uma rede profundamente supervisionada pode obter bons resultados com poucas imagens de treinamento demonstra um grande potencial, principalmente para aplicações atuais que ainda não possui grandes bancos de imagens.

## 3.2 Aprendizado Profundo em Grafos

Na subseção 3.1 foram apresentados trabalhos sobre as redes neurais convolucionais e profundamente supervisionadas. Todos estes trabalhos possuem algo em comum: as amostras (*patches*) da imagem para o treinamento são regulares (e.g. retângulos). Porém sabe-se que existem diversos problemas nos quais estruturas regulares não conseguem descreve-los. Alguns exemplos destes problemas são as redes de relacionamentos, redes de comunicação de sensores, redes funcionais em imagens cerebral, redes reguladoras em genética e superfícies gradeadas em computação gráfica [Bronstein et al., 2017]. Para solucionar esta questão uma nova linha de pesquisa tem surgido: aprendizagem profunda geométrica [Bronstein et al., 2017]. Esta metodologia tem como o objetivo generalizar a aprendizagem profunda para amostras regulares e não regulares. Uma das principais amostras não regulares são os grafos. Ultimamente, estudos de como utilizar grafos como estrutura das amostras no treinamento das CNNs têm obtido bastante atenção. Este problema não é trivial e possui diversos desafios como: domínio irregular, estruturas e tarefas variadas e, por fim, escalabilidade e paralelização.

- **Domínio irregular:** com a falta de um sistema comum de coordenadas ou uma es-

estrutura espacial vetorial, operações matemáticas se tornam mais complicadas nos grafos [Shuman et al., 2013]. Estas operações matemáticas são importantes para a convolução e o *pooling*, que são as duas principais etapas das redes neurais convolucionais.

- **Estruturas e tarefas variadas:** a dificuldade está na grande variedade de tipos de grafos e tarefas. Existem diversas estruturas como grafos com peso ou sem peso, grafos homogêneos ou heterogêneos, etc. Além disso, diversas tarefas como classificação de nós e predição de ligações dos nós, que são tarefas focadas nos nós do grafo. Mas existem também tarefas que focam na estrutura do grafo como classificação do grafo e geração de grafos.
- **Escalabilidade e paralelização:** atualmente existem problemas que podem facilmente atingir grafos com milhões de nós e conexões, como redes sociais ou redes de comércio eletrônico [Zang et al., 2016]. Por isso se torna necessário um cuidado maior com a escalabilidade e paralelização e a utilização de algoritmos capazes de trabalhar com estes tipos de grafos.

Recentemente, trabalhos da literatura têm obtido avanços para a utilização de grafos em redes neurais convolucionais. Os primeiros trabalhos usam redes neurais recursivas para processar os dados representados no grafo como grafos acíclicos [Frasconi et al., 1998, Sperduti e Starita, 1997]. Em seguida, [Gori et al., 2005] e [Scarselli et al., 2009] propuseram as redes neurais de grafos (*Graph Neural Networks* - GNNs) que é uma generalização das redes neurais recursivas dos trabalhos anteriores, porém com a capacidade de resolver problemas que utilizam grafos diretamente. As GNNs consistem em um processo iterativo que propaga as características dos nós até o seu estado de equilíbrio. Depois de atingido este estado, uma rede neural é introduzida para produzir a classe de cada nó com base em seu estado. Esta ideia foi aprimorada por [Li et al., 2015], que utilizaram unidades recorrentes bloqueadas (*gated recurrent unit* - [Cho et al., 2014]) na etapa de propagação do nó.

Atualmente existe um grande interesse em generalizar as convoluções em grafos. Os métodos

que resolvem esse problema são geralmente divididos em abordagem espectral e abordagem espacial. Na abordagem espectral o grafo tem uma representação no espectro e são em geral aplicados para classificação de nós. No trabalho de [Bruna et al., 2013], a convolução é definida no domínio de Fourier, calculando os autovetores e autovalores do grafo Laplaciano. Entretanto, esse método resulta em cálculos potencialmente intensos e filtros não espacialmente localizados. Estes problemas foram abordados por trabalhos posteriores, como [Henaff et al., 2015] que utilizaram a parametrização dos filtros espectrais com um coeficiente suave, o que pode torna-los espacialmente localizados. Para não calcular os autovetores e assim diminuir os cálculos, [Defferrard et al., 2016] propuseram aproximar os filtros espectrais através da expansão de Chebychev no grafo Laplaciano, produzindo também filtros espacialmente localizados. Já [Kipf e Welling, 2016] simplificaram o método anterior para restringir os filtros em uma vizinhança de um passo ao redor de cada nó. Nota-se que em todos estes trabalhos, os filtros aprendidos dependem da decomposição Laplaciana relacionada com a estrutura do grafo. Portanto, o modelo treinado em um determinado grafo não pode ser aplicado a grafos com estruturas diferentes.

Na abordagem espacial a operação de convolução é realizada diretamente no nó. O maior problema desta abordagem é o fato de definir um operador que consiga executar uma convolução com diferentes tamanhos de vizinhança e que consiga manter a propriedade de compartilhamento de pesos das redes neurais convolucionais. [Duvenaud et al., 2015] propuseram aprender uma matriz de peso específica para cada grau de nó. Já [Atwood e Towsley, 2015] usam uma matriz de transição para definir a vizinhança da convolução enquanto aprendem os pesos para cada entrada e grau de vizinhança. Outra proposta foi extrair e normalizar a vizinhança através de um número de nós fixos [Niepert et al., 2016]. [Monti et al., 2017] apresentaram uma abordagem espacial que tenta unificar a generalização das arquiteturas de redes neurais convolucionais para problemas com grafos. Um trabalho mais recente desenvolvido por [Hamilton et al., 2017] apresentaram um método que os autores nomearam de *GraphSAGE*. Este método calcula as representações dos nós de uma maneira indutiva, funcionando para

---

uma vizinhança com tamanho fixo de cada nó. Depois desta representação, uma operação de soma específica (e.g. média de todos os vetores de características dos vizinhos) é realizada para cada nó. Esta abordagem têm recebido grande atenção devido ao grande desempenho que obteve em problemas de larga escala.

## 4 Metodologia

Para alcançar os objetivos propostos, este projeto será desenvolvido obedecendo as seguintes fases:

### **Meta 1: Criação de um banco de imagens de bovinos e de carcaças**

Inicialmente acontecerá visitas em fazendas e frigoríficos para captura imagens de bovinos e carcaças. Pretende-se que estas imagens sejam tanto em ambientes controlados quanto em ambientes não-controlados. O banco de imagens será composto com diversas imagens de gados em diferentes ambientes: no curral ou soltos em um cercado. Já existem dois bancos que estão sendo produzidos, um banco de imagens de carcaças e um com os animais soltos em um cercado. O bando de imagens de carcaça possui 226 imagens com dimensão  $3264 \times 2448$  que foram capturadas em um frigorífico da região de Mato Grosso do Sul<sup>1</sup> utilizando um celular Samsung Galaxy S7. As imagens possuem diferentes tipos de escala, iluminação e ângulos. A Figura 1 apresenta 2 exemplos em que podemos observar que a captura foi em um ambiente interno do frigorífico, mas não controlado. O segundo banco de imagens que está sendo construído consiste em animais soltos em um ambiente não controlado e será utilizado para o problema de estimar o peso do animal. Este banco de imagens é composto por 154 imagens com dimensão de 4032 por 3024 pixels com diferentes escalas, iluminações e ângulos, o que a torna muito complexa. A Figura 2 apresenta 2 exemplos que demonstram a complexidade desta banco.

### **Meta 2: Aprofundamento da revisão bibliográfica**

---

<sup>1</sup>O nome do frigorífico foi omitido por questões legais.



Figura 1: Exemplos de duas imagens presentes no banco de imagens da carcaça que está sendo construído. Partes da imagem estão borrada devido a questões legais.



Figura 2: Exemplos de duas imagens presentes no banco de imagens de animais soltos em um cercado.

Nesta etapa a revisão bibliográfica será aprofundada com aplicações da pecuária de precisão em visão computacional com foco em projetos que utilizaram redes neurais convolucionais profundamente supervisionada para classificar carcaças ou para tentar estimar o peso do animal. Através de consultas aos principais portais de periódicos, como IEEE Xplore, ACM DL, Science Direct e Scopus, serão identificados os trabalhos correlatos relacionados as abordagens

---

usadas neste trabalho: rede neural convolucional profundamente supervisionada, superpixel e grafos. Nesta etapa do projeto também serão estudadas as formas de combinações entre estas abordagens propostas na literatura.

### **Meta 3: Desenvolvimento e implementação de métodos para os problemas abordados e validação do método proposto**

Após o estudo de métodos da literatura, alguns destes métodos serão escolhidos para implementação afim de obter um estudo mais aprofundado. Então será possível desenvolver métodos profundamente supervisionado e/ou cuja a convolução ocorra através de estruturas não regulares. Os métodos profundamente supervisionados terão como base o trabalho de [Bazi e Melgani, 2018], porém com algumas diferenças:

- **Camadas convolucionais com SVM multi-classes:** a proposta original funciona somente com duas classes, a ideia é estender a capacidade de classificação da rede para problemas multi-classes. Para isso, os SVM das camadas serão implementados com estratégias multi-classe, tais como 1-vs-1 e 1-vs-todos.
- **Treinar as camadas convolucionais com SVM de forma dependente:** no trabalho original, as camadas convolucionais com SVM são treinadas de forma independente, ou seja, o treinamento dos SVMs das camadas posteriores só é iniciado com a finalização do treinamento da sua camada anterior respectivamente. Por exemplo, o treinamento dos SVMs da 5<sup>a</sup> camada convolucional só inicia se os da 4<sup>a</sup> camada estiver finalizado, e assim sucessivamente. Além disso, após terminar o treinamento das camadas, os pesos dos SVMs são fixados para treinar a próxima camada. Assim, pretende-se estudar uma forma de treinar as camadas convolucionais com SVM de forma dependente, assim como acontece no *backpropagation* das redes neurais convolucionais tradicionais.
- **Estender para segmentação:** assim como a proposta de [Badrinarayanan et al., 2017] que propuseram uma rede neural convolucional especializada em segmentação, a ideia é

---

propor uma rede neural convolucional profundamente supervisionada que seja utilizada para segmentação. A principal diferença será o uso das camadas convolucionais com SVM, ao invés das camadas convolucionais tradicionais.

- **Inception:** esta é uma técnica que tem atraído atenção atualmente. Este método utiliza a ideia de obter uma camada que possua a capacidade de desempenhar convoluções com diferentes tamanhos de filtros. Almeja-se estender esta técnica para as redes neurais convolucionais profundamente supervisionadas com SVM.

Para os métodos cuja convolução ocorre através de estruturas não regulares, que também serão avaliados na aprendizagem profundamente supervisionada, pretende-se propor as seguintes ideias:

- **Usar extratores de características:** a ideia é utilizar extratores de características tradicionais em visão computacional para representar uma imagem através de um grafo. Um extrator que será utilizado é o SLIC Superpixel [Achanta et al., 2012], que tem como objetivo dividir uma imagem em regiões com propriedades similares como cor, textura e brilho. Estas regiões geralmente não são regulares, portanto a ideia é utilizar os superpixels gerados pelo SLIC como nós do grafo. Dois nós serão conectados se as regiões dos seus superpixels correspondentes forem vizinhas na imagem. Dado o grafo que representa a imagem de entrada, convoluções com estruturas não regulares serão aplicadas. A vantagem é que o grafo apresenta informações adicionais, tais como as regiões e suas conexões na imagem, que podem auxiliar no reconhecimento da imagem.
- **Aprendizado residual:** uma das redes neurais convolucionais de maior sucesso na literatura é a ResNet, proposta por [Ren et al., 2015]. A ideia dos autores foi utilizar o aprendizado residual para treinar CNNs cada vez mais profundas. A ideia foi motivada pelo fenômeno conhecido como degradação, onde o aprendizado das camadas mais profundas acaba ficando saturado. Por isso, os autores tiveram a ideia de somar ponto a

---

ponto a entrada de uma camada convolucional à saída de outra usando a Equação 1. Assim, pretende-se propor o aprendizado residual para redes neurais convolucionais cuja convolução seja utilizando estruturas não regulares. Dessa forma, o grafo de entrada de uma camada convolucional será somado ao grafo da saída, por exemplo, somando os vetores de características de cada nó.

$$V_s = F(V_e) + V_e, \quad (1)$$

onde  $V_e$  e  $V_s$  são a entrada e a saída da camada residual considerada,  $F(V_e)$  é o volume de saída após a aplicação das duas camadas de convolução.

- **Pooling com grafos:** sabe-se que a estrutura utilizada no *pooling* nas redes neurais convolucionais tradicionais também são regulares. Neste caso, a ideia é utilizar estruturas não regulares para realizar a etapa de *pooling* na rede. Uma das tentativas será representar um conjunto de nós conectados entre si com um único nó aplicando-se uma função nas características.

Após a implementação, os métodos serão adicionados ao *software* Pynovisão. O Pynovisão é um *software* desenvolvido pelo grupo INOVISÃO e tem como base o uso do algoritmo de visão computacional para segmentar e classificar imagens. Também será nesta parte do projeto que os métodos proposto serão validados para diferentes tipos de imagens. *A priori*, como métricas serão utilizadas a porcentagem de classificação correta (PCC), medida-f, precisão e revocação. Para amostragem será adotada a validação cruzada de 10 dobras com 10 repetições (*10-folders cross-validation*). Para comparar o método proposto com métodos da literatura, será utilizada a análise de variância (ANOVA), que identificará se existe diferença estatísticas entre os métodos. Por fim, um pós-teste será realizado para comparar estatisticamente os métodos.

**Meta 4: Aplicação dos métodos propostos para os problemas de pecuária de precisão**

---

Nesta etapa os métodos proposto serão aplicados para a classificação de carcaça e estimativa de peso. O processo de tipificação consiste em dividir a carcaça em classes de qualidade, sendo os principais critérios para esta divisão sua cobertura de gordura e conformação. Esta tipificação é importante porque permite uma definição da classe da carcaça, afetando o seu valor para o mercado interno e para o mercado externo. Porém, este processo geralmente é feito por um especialista, o que pode tornar a tipificação subjetiva. Nesse problemas, os métodos serão usados para classificar uma imagem da carcaça em uma das cinco classes de qualidade atualmente utilizadas no Brasil.

O segundo problema abordado é a estimativa do peso de bovinos através de imagens em ambientes não controlado, considerado como um problema bem desafiador. A ideia é tornar este processo mais rápido e sem estresse para o animal, facilitando o controle do rebanho para o produtor. Além disso, alguns produtores menores não possuem uma balança para obter o peso do animal. Esta estimativa do peso não é necessária somente no momento da venda para um frigorífico, mas também para notar problemas de saúde do animal, perceber se a aplicação de um medicamento está surtindo efeito, bem como constatar se a dieta do animal está sendo bem realizada. Portanto, nesta etapa os métodos serão aplicados nestes dois problemas tão importantes na pecuária de precisão.

### **Meta 5: Produção de textos**

Nesta etapa os resultados obtidos serão divulgados por meio de artigos científicos e palestras. Revistas, eventos nacionais e internacionais apropriados serão selecionados para submissão dos resultados. Além disto, serão produzidos os textos da qualificação e da tese nesta etapa.

## **5 Atividades e Cronograma de Execução**

As metas descritas na metodologia serão executadas seguindo o cronograma e as atividades da Tabela 1.

Metas	Atividades	Tempo (trimestral)							
		Ano 1		Ano 2		Ano 3		Ano 4	
		1°	2°	1°	2°	1°	2°	1°	2°
Meta 1	1.1 - Definir os critérios/ambientes para a captura de imagens	•							
	1.2 - Capturar as imagens em diferentes ambientes e condições	•							
	1.3 - Criação dos banco de imagens	•	•						
	1.4 - Pré-processamento das imagens		•						
Meta 2	2.1 - Revisão bibliográfica sobre visão computacional e pecuária de precisão	•	•						
	2.2 - Revisão bibliográfica sobre redes neurais convolucionais		•	•					
	2.4 - Revisão bibliográfica sobre aprendizagem profunda em grafos		•	•					
Meta 3	3.1 - Selecionar métodos estudados na revisão para serem implementados			•					
	3.2 - Propor os métodos de redes neurais convolucionais profundamente supervisionados			•	•	•	•		
	3.2 - Propor os métodos de aprendizado profundo em grafos			•	•	•	•		
	3.4 - Testar e avaliar os métodos implementados em diferentes ambientes				•	•	•		
Meta 4	4.1 - Ajustar os métodos propostos					•	•	•	
	4.2 - Aplicar os métodos em problemas da pecuária de precisão						•	•	
	4.3 - Produção dos indicadores de desempenho (procentagem de classificação correta e outras medidas)							•	•
Meta 5	5.1 - Produção do texto da qualificação			•	•				
	5.2 - Produção e divulgação dos resultados por meio de artigos científicos e palestras			•	•	•	•	•	•
	5.3 - Produção do texto da tese							•	•

Tabela 1: Cronograma de execução das metas e atividades.

## 6 Resultados Esperados, Produtos e Avanços

Este projeto contribuirá para o desenvolvimento teórico e prático das áreas de visão computacional e pecuária de precisão. Espera-se propor um método robusto e eficaz de redes neurais convolucionais com estruturas não regulares na convolução. Como estrutura não regular será utilizado o grafo, já que diversos extratores de características da visão computacional, tais como superpixels, resultam em um grafo. Almeja-se também propor métodos profundamente super-

visionados com arquiteturas que possuam mais camadas, uma vez que [Bazi e Melgani, 2018] apresentaram resultados com 5 camadas convolucionais. Acredita-se que o incremento no número de camadas resultará em bons resultados caso essas camadas sejam treinadas com dependência, uma vez que [Bazi e Melgani, 2018] treinaram cada camada independentemente.

A constatação da eficácia destes métodos propostos contribuirá para a solução de problemas importantes da área da pecuária de precisão. Estes métodos aplicados na estimativa do peso do animal podem auxiliar o produtor e especialistas veterinários na tomada de decisão para aplicar medidas preventivas para o melhora da saúde do animal. Espera-se também aplicar os métodos na tipificação de carcaça com o intuito de retirar a parte subjetiva deste processo. Sabe-se que o auxílio ao produtor e aos especialistas nestes dois problemas terão um grande impacto no estado de Mato Grosso do Sul, visto que a pecuária é uma área muito forte economicamente nesta região. Pretende-se também contribuir com a criação de dois bancos de imagens anotados: carcaça e de bovinos soltos no curral. Além disso, uma biblioteca com os códigos serão disponibilizadas através do *software* Pynovisão. Por fim, estima-se que o projeto resulte publicações em conferências e revistas nacionais e internacionais.

## Referências

- [Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., e Süsstrunk, S. (2012). SLIC Superpixels Compared to State-of-the-art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274 – 2282.
- [Associação Brasileira Indústrias Exportadoras Carnes, 2017] Associação Brasileira Indústrias Exportadoras Carnes (2017). Perfil da pecuária no brasil.
- [Atwood e Towsley, 2015] Atwood, J. e Towsley, D. (2015). Diffusion-convolutional neural networks.

- [Badrinarayanan et al., 2017] Badrinarayanan, V., Kendall, A., e Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495.
- [Bazi e Melgani, 2018] Bazi, Y. e Melgani, F. (2018). Convolutional svm networks for object detection in uav imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6):3107–3118.
- [Bronstein et al., 2017] Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., e Vandergheynst, P. (2017). Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42.
- [Bruna et al., 2013] Bruna, J., Zaremba, W., Szlam, A., e LeCun, Y. (2013). Spectral networks and locally connected networks on graphs.
- [Cho et al., 2014] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., e Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation.
- [Ciresan et al., 2012] Ciresan, D., Meier, U., e Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. *2012 IEEE Conference on Computer Vision and Pattern Recognition*.
- [Dahl et al., 2012] Dahl, G. E., Yu, D., Deng, L., e Acero, A. (2012). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):30–42.
- [Defferrard et al., 2016] Defferrard, M., Bresson, X., e Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering.

- [Duvenaud et al., 2015] Duvenaud, D., Maclaurin, D., Aguilera-Iparraguirre, J., Gómez-Bombarelli, R., Hirzel, T., Aspuru-Guzik, A., e Adams, R. P. (2015). Convolutional networks on graphs for learning molecular fingerprints.
- [Eigen et al., 2013] Eigen, D., Rolfe, J., Fergus, R., e LeCun, Y. (2013). Understanding deep architectures using a recursive convolutional network.
- [Frasconi et al., 1998] Frasconi, P., Gori, M., e Sperduti, A. (1998). A general framework for adaptive processing of data structures. *IEEE Transactions on Neural Networks*, 9(5):768–786.
- [Frost et al., 1997] Frost, A., Schofield, C., Beulah, S., Mottram, T., Lines, J., e Wathes, C. (1997). A review of livestock monitoring and the need for integrated systems. *Computers and Electronics in Agriculture*, 17(2):139 – 159.
- [Gori et al., 2005] Gori, M., Monfardini, G., e Scarselli, F. (2005). A new model for learning in graph domains. Em *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, volume 2, páginas 729–734 vol. 2.
- [Graves et al., 2013] Graves, A., Mohamed, A., e Hinton, G. (2013). Speech recognition with deep recurrent neural networks. Em *IEEE International Conference on Acoustics, Speech and Signal Processing*, páginas 6645–6649.
- [Hamilton et al., 2017] Hamilton, W. L., Ying, R., e Leskovec, J. (2017). Inductive representation learning on large graphs.
- [Henaff et al., 2015] Henaff, M., Bruna, J., e LeCun, Y. (2015). Deep convolutional networks on graph-structured data.
- [Hinton et al., 2012] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., e Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors.

- [Karpathy et al., 2014] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., e Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 1725–1732.
- [Kipf e Welling, 2016] Kipf, T. N. e Welling, M. (2016). Semi-supervised classification with graph convolutional networks.
- [Kist, 2017] Kist, B. B. (2017). Anuário brasileiro do gado de corte. *Editora Gazeta Santa Cruz Ltda.*
- [Krizhevsky et al., 2012a] Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012a). Imagenet classification with deep convolutional neural networks. Em Pereira, F., Burges, C. J. C., Bottou, L., e Weinberger, K. Q., editores, *Advances in Neural Information Processing Systems 25*, páginas 1097–1105. Curran Associates, Inc.
- [Krizhevsky et al., 2012b] Krizhevsky, A., Sutskever, I., e Hinton, G. E. (2012b). Imagenet classification with deep convolutional neural networks. Em Pereira, F., Burges, C. J. C., Bottou, L., e Weinberger, K. Q., editores, *Advances in Neural Information Processing Systems 25*, páginas 1097–1105. Curran Associates, Inc.
- [LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., e Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551.
- [Lee et al., 2015] Lee, C.-Y., Xie, S., Gallagher, P. W., Zhang, Z., e Tu, Z. (2015). Deeply-supervised nets. *CoRR*, abs/1409.5185.
- [Li et al., 2015] Li, Y., Tarlow, D., Brockschmidt, M., e Zemel, R. (2015). Gated graph sequence neural networks.

- [Luo et al., 2012] Luo, P., Wang, X., e Tang, X. (2012). Hierarchical face parsing via deep learning. Em *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 2480–2487.
- [Monti et al., 2017] Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., e Bronstein, M. M. (2017). Geometric deep learning on graphs and manifolds using mixture model cnns. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [Mostafa et al., 2018] Mostafa, H., Ramesh, V., e Cauwenberghs, G. (2018). Deep supervised learning using local errors. *Frontiers in Neuroscience*, 12.
- [Niepert et al., 2016] Niepert, M., Ahmed, M., e Kutzkov, K. (2016). Learning convolutional neural networks for graphs.
- [Ren et al., 2015] Ren, S., He, K., Girshick, R., e Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Em Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., e Garnett, R., editores, *Advances in Neural Information Processing Systems 28*, páginas 91–99. Curran Associates, Inc.
- [Russakovsky et al., 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., e Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252.
- [Scarselli et al., 2009] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., e Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80.
- [Shuman et al., 2013] Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., e Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98.

- [Sperduti e Starita, 1997] Sperduti, A. e Starita, A. (1997). Supervised neural networks for the classification of structures. *IEEE Transactions on Neural Networks*, 8(3):714–735.
- [Sun et al., 2013] Sun, Y., Wang, X., e Tang, X. (2013). Deep convolutional network cascade for facial point detection. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 3476–3483.
- [Toshev e Szegedy, 2014] Toshev, A. e Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 1653–1660.
- [Wang et al., 2015] Wang, L., Lee, C.-Y., Tu, Z., e Lazebnik, S. (2015). Training deeper convolutional networks with deep supervision.
- [Zang et al., 2016] Zang, C., Cui, P., e Faloutsos, C. (2016). Beyond sigmoids: The nettide model for social network growth, and its applications. Em *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, páginas 2015–2024, New York, NY, USA. ACM.
- [Zeiler e Fergus, 2014] Zeiler, M. D. e Fergus, R. (2014). Visualizing and understanding convolutional networks. Em Fleet, D., Pajdla, T., Schiele, B., e Tuytelaars, T., editores, *Computer Vision – ECCV 2014*, páginas 818–833, Cham. Springer International Publishing.