UNIVERSIDADE FEDERAL
DE ALAGOAS

INSTITUTO DE COMPUTAÇÃO
PROCESSAMENTO DIGITAL DE IMAGENS

RELATÓRIO
TRABALHO SELECIONADO PARA O PROJETO E PESQUISA BIBLIOGRÁFICA

Nome: Lucas Mendes Massa Matrícula: 18112211
Nome: Jhonnye Gabriel de Oliveira Farias Matrícula:

TRABALHO BASE

In Defense of Classical Image Processing: Fast Depth Completion on the CPU

Neste artigo, os autores propõem uma abordagem de processamento de imagem clássico para atacar o problema de depth completion, uma importante tarefa no campo da robótica e visão de máquina. Esse problema se resume a inferir mapas de profundidade densos a partir de entradas compostas de imagens em conjunto com mapas de densidade esparsos. Isso se dá pelo fato de que sensores LIDAR conseguem apenas fornecer mapas de densidade esparsos, o que limita a performance e a aplicabilidade de diversos algoritmos de percepção.

Com os grandes avanços existentes na área da aprendizagem profunda, muitos pesquisadores têm abandonado métodos tradicionais de processamento de imagem. No entanto, o treinamento e a execução de redes neurais ainda exige um poder computacional bastante elevado. Geralmente os modelos de rede neural rodam em GPUs, as quais consomem muita energia e inviabilizam a implantação de sistemas de depth completion em sistemas embarcados. Dessa forma, os autores se propõem a mostrar que é possível obter um algoritmo com desempenho satisfatório fazendo uso apenas de métodos clássicos de processamento de imagem. No tempo em que o trabalho foi publicado, o algoritmo dos autores havia obtido o melhor desempenho no benchmarking de depth completion da base de dados KITTI. Além disso, o algoritmo roda em CPU e possui um tempo de execução menor que seus concorrentes da classe de aprendizagem profunda.

O algoritmo proposto é formado basicamente por 8 passos:

1) Inversão de Profundidade: Ao considerar os dados brutos do mapa de profundidade KITTI, pixels mais próximos assumem valores próximos a 0 m, enquanto os mais afastados tomam valores até um máximo de 80 m. Porém, pixels vazios também assumem o valor 0 m, o que impede o uso de Operações OpenCV sem modificação. Para resolver este problema, as profundidades de pixel válidas (não vazias) são invertidas .
2) Dilatação de kernel personalizada: começa-se preenchendo os pixels vazios mais próximos dos pixels válidos, pois é provável que eles compartilhem valores de profundidade próximos ao de profundidades válidas. Considerando tanto a esparsidade de pontos projetados e a estrutura das linhas de varredura do LIDAR, foi projetado um kernel personalizado para a dilatação inicial de cada pixel de profundidade válido.
3) Fechamento de Pequenos Furos: Após a etapa inicial de dilatação, muitos buracos ainda existem no mapa de profundidade. Uma vez que essas áreas não contêm valores de profundidade, considera-se a estrutura dos objetos no ambiente, pois manchas próximas de dilatação profundidades podem ser conectadas para formar as bordas dos objetos. Uma operação de fechamento morfológico, com um kernel cheio 5 × 5, é usada para fechar pequenos buracos no mapa de profundidade.
4) Preenchimento de furos pequenos: alguns furos de tamanho pequeno a médio no mapa de profundidade não são preenchidos pelas duas primeiras dilatações. Para preencher esses buracos, uma máscara de pixels vazios é calculada, seguida por uma dilatação total com um kernel 7 × 7. Operação. Esta operação resulta apenas nos pixels vazios sendo preenchidos, mantendo pixels válidos que foram anteriormente calculados inalterados.
5) Extensão para o topo do quadro: para considerar objetos altos como árvores, postes e edifícios que se estendem acima do topo dos pontos do LIDAR, o valor superior ao longo de cada coluna é extrapolado para o topo da imagem, proporcionando um mapa de profundidade mais denso na saída.
6) Preenchimento de furos grandes: A etapa final de preenchimento cuida de buracos no mapa de profundidade que ainda não estão totalmente preenchidos. Como essas áreas não contêm pontos e nenhum dado de imagem é usado, os valores de profundidade para esses pixels são extrapolados de valores próximos. Uma operação de dilatação com um 31x31 cheio kernel é usada.
7) Borramento Mediano e Gaussiano: Após aplicar os passos anteriores, tem-se um mapa de profundidade denso. No entanto, existem outliers nesse mapa de profundidade como um subproduto das operações de dilatação. Para remover esses outliers, faz-se uso de borramentos.
8) Inversão de profundidade: O passo final do algoritmo é reverter para a codificação de profundidade original.

ARTIGOS DE REVISTA

A comparative review of plausible hole filling strategies in the context of scene depth image completion

Survey que visa fornecer uma visão geral do estado da arte a respeito da síntese da profundidade, preenchimento da estrutura de profundidade subjacente e do relevo de texturas. A importância desse tópico é justificada pelo fato de que, em sistemas de captura de cena 3D, inúmeros desafios relacionados a alcançar uma cobertura total para a estimativa de profundidade de cena ainda permanecem em aberto. Essa estimativa é crucial para qualquer sistema de detecção 3D moderno. No que diz respeito a depth completion, o trabalho trata de temas como taxonomia dos avanços existentes, aspectos de formulação de problemas, consistência espacial e continuidade temporal. Como se trata de um trabalho um pouco mais antigo, ainda não é dada atenção a métodos de aprendizagem profunda aplicados a essa problemática, tendo enfoque em métodos mais tradicionais. Dentre as técnicas relatadas no trabalho pode-se citar difusão anisotrópica, minimização de energia, preenchimento baseado em exemplos, conclusão de matriz, filtragem, interpolação e extrapolação. Por fim, é dito que pesquisas futuras precisam considerar explicitamente a eficiência computacional e que também é altamente provável que seja possível explorar aspectos temporais de um “fluxo de profundidade” ao vivo, o que teria um grande suporte de métodos de machine learning.

Depth map artefacts reduction: a review

Survey a respeito de métodos de redução de artefatos em mapas de profundidade, desde mono até multi-view, via dimensão espacial ou temporal, de local para forma global, utilizando tanto processamento de sinais quanto métodos baseados em aprendizagem. Novamente é dada a justificativa de que os mapas de profundidade são de suma importância para muitas aplicações visuais, já que os mesmos representam informações de posicionamento dos objetos em uma cena tridimensional. Para o tema de depth completion com dados de sensores LIDAR é dado destaque especial para técnicas de deep learning, mais especificamente redes neurais convolucionais (CNNs). Algumas das redes mencionadas fazem uso de uma máscara de validação adicional. Essas máscaras de validação são usadas para identificar os pixels ausentes através da atribuição de zero ao valor dos mesmos. Apesar de redes recorrentes serem citadas dentro de outras seções, é dito que, à época em que artigo foi escrito, as mesmas ainda não tinham sido devidamente exploradas para o processamento de mapas de densidade, sendo esse um problema em aberto. Dessa forma, o estudo de outras arquiteturas é proposto como trabalho futuro, além da investigação de como se construir redes neurais mais leves e que possam ter aplicações práticas.

A Survey on Deep Learning Techniques for Stereo-based Depth Estimation

Survey sobre aprendizado profundo para estimativa de profundidade baseada em imagens estéreo, onde foram resumidos os pipelines mais usados e discutidos seus benefícios e limitações. Apesar da grande quantidade de pesquisas, técnicas tradicionais ainda sofrem com a presença de áreas altamente texturizadas, grandes regiões uniformes e oclusões. Dessa forma, motivados por seu crescente sucesso na solução de vários problemas de visão 2D e 3D, o aprendizado profundo para estimativa de profundidade baseada em estéreo tem atraído grande interesse da comunidade. No que diz respeito à tarefa específica de depth completion, é mencionado que Redes Convolucionais de Propagação Espacial, são particularmente adequadas, pois implementam um processo de difusão anisotrópica.

ARTIGOS DE CONFERÊNCIA

Sparsity Invariant CNNs

O paper considerou redes neurais convolucionais operando em entradas esparsas com uma aplicação para amostragem de profundidade a partir de dados de varredura a laser esparsos. Primeiramente, mostrando que CNNs tradicionais performam de forma precária quando aplicados em dados esparsos, mesmo quando a localização do dado faltante é providenciado para a rede

Para superar esse problema, foi proposto uma simples, mas ainda efetiva, camada de convolução esparsa que explicitamente considera a localização do dado faltante durante a operação de convolução.

Os benefícios da arquitetura de rede proposta são demonstrados em experimentos sintético e reais com relação a várias abordagens de linha de base. Comparado a linhas de base densas, a proposta generaliza muito bem novos conjuntos de dados e é invariável ao nível de esparsidade nos mesmos. A avaliação foi derivada de um novo conjunto de dados do benckmark KITTI, que compreende imagens RGBs anotadas com profundidade de 93k. Sendo assim o novo módulo de convolução esparso proposto resultou em melhor desempenho enquanto generalizando para novos domínios ou níveis de esparsidade.

Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera

O preenchimento de profundidade enfrenta três principais: O padrão irregularmente espaçado nas entradas de profundidade esparsa, a dificuldade de lidar com as várias modalidades de sensores (ex: diferentes tipos de imagens), bem como a falta de densidade, rótulos de profundidade do chão verdadeiros a nível de pixel, para o treinamento. O presente trabalho pontua todos esses desafios, especialmente no desenvolvimento de um modelo de regressão profundo para aprender um mapeamento direto da entrada de profundidade esparsa (e imagens coloridas) para a previsão de profundidade densa.

Também foi proposto um framework de treinamento auto-supervisionado que requer apenas sequenciamento de cores e imagens de profundidade esparsa, sem necessitar de rótulos de profundidade densa.

Os experimentos demonstraram que o framework superou um número de soluções treinadas existentes com anotações semi-densas. Além disso, quando treinado com essas mesmas anotações, a rede atinge a precisão do estado da arte e possui uma abordagem vencedora na base de dados de preenchimento de profundidade KITTI, no benchmarking de no tempo que foi submetido. (2019)

LIDAR and Monocular Camera Fusion: On-road Depth Completion for Autonomous Driving

LIDAR e câmeras RGBs são comumente usadas para sensores em veículos autônomos, mas como ambos, o LIDAR e Câmeras RGBs possuem limitações: LIDAR provê acurácia na profundidade, mas é esparso na resolução vertical e horizontal; Imagens RGBs provém textura densa, mas uma falta de informação de profundidade.

O trabalho ataca nesse problema para fundir ambos os componentes em uma rede neural densa, que completa um mapa de profundidade em pixels mais denso. A arquitetura propõe reconstruir o mapa de profundidade em pixels, aproveitando os recursos de cores densas e os recursos espaciais 3D esparsos.

O método foi avaliado nos conjuntos de dados de odometria interna de grande escala NYUdepthV2 e KITTI, superando o estado da arte nos métodos de fusão profunda e de única imagem RGB. O método também foi avaliado no conjunto de dados KITTI de baixa resolução, que sintetiza a fusão planar de imagens LIDAR e RGB.

Depth completion from sparse lidar data with depth normal constraints

A maioria dos métodos competitivos atuais treina diretamente uma rede para aprender a mapear de entradas de profundidade esparsas para mapas de profundidade densos, que por sua vez, possuem dificuldades em utilizar as restrições da geometria tridimensional e também problemas ao lidar com ruídos do sensor.

Nesse artigo, para regularizar e melhorar a robustez do preenchimento de profundidade contra o ruído, eles propõem um framework da rede neural convolucional unificada que:

1) Modela as restrições geométricas entre profundidade e superfície normal em um módulo de difusão
2) Prevê a confiança de medições do lidar esparso visando mitigar o impacto do ruído. Essa previsão está implementada na estrutura do seu decodificador e encodificador, prevendo superfícies normais, profundidade grosseiras e a confiança das entradas do lidar. Esses dados ainda são refinados no módulo de difusão para a obtenção dos resultados finais.

Foram realizados experimentos extensivos na base de dados KITTI de preenchimento de profundidade e na base de dados NYU-Depth-V2. O método demonstrou alcançar um desempenho equiparado à um estado da arte. Outros estudos também demonstram que os componentes propostos no método possuem capacidade de generalização e estabilidade.

Deep Adaptive LiDAR: End-to-end Optimization of Sampling and Depth Completion at Low Sampling Rates

Os sistemas de LiDAR atuais são limitados na sua habilidade de capturar nuvem de pontos tridimensionais densas. Para contornar esse problema, algoritmos de preenchimento de profundidade baseados em deep learning estão sendo desenvolvidos para pintar a profundidade ausente guiada por uma imagem RGB. Porém esses métodos falham em situações de baixa taxa de amostragem.

O presente trabalho propõe um esquema de amostragem adaptativa para sistemas de LiDAR que demonstram desempenho de estado da arte para preenchimento de profundidade à baixas taxas amostragem. O sistema é diferenciável, permitindo que a amostragem de profundidade esparsa e os componentes de pintura de profundidade sejam treinados e ponta a ponta com uma tarefa *upstream.*

From Depth What Can You See? Depth Completion via Auxiliary Image Reconstruction

Os métodos somente de profundidade existentes usam apenas profundidade esparsa como entrada. No entanto, esses métodos podem falhar em recuperar limites semânticos consistentes ou objetos pequenos/finos devido:

1) a natureza esparsa dos pontos de profundidade
2) a falta de imagens para fornecer pistas semânticas

O presente trabalho dá continuidade a essa linha de pesquisa e visa superar as dificuldades acima. O que traz o design único para esse modelo de preenchimento de profundidade é que produz simultaneamente uma imagem reconstruída e um mapa de profundidade denso. Especialmente, é formulado a reconstrução de imagens a partir de profundidade esparsas como uma tarefa auxiliar durante o treinamento supervisionado com imagens em escala de cinza não-rotuladas. Esse

mesmo design permite que, a rede de preenchimento de profundidade, aprenda recursos de imagens complementares, que ajudam no melhor entendimento das estruturas dos objetos.

O método foi avaliado na base de dados KITTI de preenchimento de profundidade em seu benchmark e mostrou que o preenchimento pode ser significativamente aprimorado através do auxílio supervisionado de reconstrução de imagens. O mesmo superou os métodos que usam somente profundidade e é também efetivo para cenários internos como no NUYv2.

Depth Completion via Inductive Fusion of Planar LIDAR and Monocular Camera

Um LIDAR moderno de alta definição é muito caro para veículos comerciais autônomos e pequenos robôs. Uma solução plausível para esse problema é a fusão de um lidar planar com imagens RGBs para prover um nível similar ao da capacidade de percepção. E mesmo os métodos em estado da arte, que provém uma abordagem para prever a informação de profundidade de entradas de sensores limitados, geralmente eles fazem uma concatenação dos recursos de um lidar esparso e de RGB densos através de uma arquitetura de fusão ponta-a-ponta.

O presente trabalho introduz um bloco de fusão tardia indutiva, que funde melhor diferentes modalidades de sensores inspirados em um modelo de probabilidade. A rede de demonstração e agregação proposta propaga as características mistas de contexto e profundidade para a rede de previsão e serve como um conhecimento prévio do preenchimento de profundidade. Esse bloco de fusão tardia usa as características de contexto denso para orientar a previsão de profundidade baseado em demonstrações por características de profundidade esparsas.

Em adição, para a avaliação, o método foi submetido ao *benchmark* de bases de dados de preenchimento de profundidade incluindo NYUDepthV2 e KITTI, também foi testado em uma simulação de base de dados de LIDAR planar, mostrando resultados promissores comparados para abordagens anteriores em ambos os *benchmarks* e conjuntos de dados simulados com variadas densidades 3D.

Radar-Camera Pixel Depth Association for Depth Completion

Embora os dados de radar e vídeo possam ser fundidos prontamente no nível de detecção, fundi-los no nível do pixel é potencialmente mais benéfico. Isso também é mais desafiador em parte devido à esparsidade do radar, mas também porque os feixes de radar automotivos são muito mais largos do que um pixel típico combinado com uma grande linha de base entre a câmera e o radar, o que resulta em má associação entre pixels de radar e pixel de cor. Uma consequência é que os métodos de preenchimento de profundidade projetados para LiDAR e vídeo se saem mal para radar e vídeo.

O trabalho propõe um estágio de associação radar-pixel que aprende um mapeamento do radar retorna aos pixels. Esse mapeamento também serve para adensar os retornos do radar. Usando esse método como primeiro estágio, seguindo de um outro método tradicional de preenchimento de profundidade, eles foram

capazer de alcançar o preenchimento guiado por imagem com radar e vídeo.

Foi demonstrado desempenho superior para somente radar e câmeras na base de dados nuScenes. O experimento mostra também que preenchimento de profundidade usando MER obtém acurácia elevada em relação ao radar bruto.

Grayscale And Normal Guided Depth Completion With A Low-Cost Lidar

Neste artigo, é apresentado o DenseLivox, um conjunto de dados com profundidade densa e precisa como verdade no terreno. Até a data da publicação foi o primeiro conjunto de dados com verdade do solo densa projetado para o preenchimento de profundidade LiDAR usando um LiDAR de baixo custo.

Além disso, foi desenvolvido uma rede de aprendizado multitarefa simples, mas eficaz, para resolver o problema da conclusão em profundidade.

Comparado com os trabalhos da literatura, a singularidade do modelo é que ele completa um mapa de profundidade, um mapa normal e uma imagem em tons de cinza simultaneamente. Para abordar a área com ruídos pesados, foi usada a perda de Huber modificada para suavizar o efeito desses valores discrepantes.

Avaliamos nosso método no DenseLivox e mostrado que a precisão é muito melhorada com a escala de cinza e a orientação normal. Nosso método supera outros métodos somente de profundidade e é comparável aos métodos que usam RGB e profundidade como entrada.

UAMD-Net: A Unified Adaptive Multimodal Neural Network for Dense Depth Completion

A previsão de profundidade é um problema crítico em aplicações de robótica, especialmente a condução autônoma. No entanto, o primeiro geralmente sofre de *overfitting* durante a construção do volume de custo, e o segundo tem uma generalização limitada devido à falta de restrição geométrica.

Para resolver esses problemas, foi proposto no trabalho uma nova rede neural multimodal, chamada UAMD-Net, para preenchimento de profundidade densa com base na fusão de correspondência estéreo binocular e a restrição fraca das nuvens de pontos esparsos. Especificamente, as nuvens de pontos esparsos são convertidas em mapa de profundidade esparsa e enviadas para o codificador de recurso multimodal (MFE) com imagem binocular, construindo um volume de custo multimodal. Em seguida, será processado pelo agregador de recursos multimodal (MFA) e pela camada de regressão de profundidade.

Além disso, os métodos multimodais existentes ignoram o problema da dependência modal, ou seja, a rede não funcionará quando uma determinada entrada modal apresentar um problema. Portanto, essa nova estratégia de treinamento chamada

Modal-dropout que permite que a rede seja treinada de forma adaptativa com múltiplas entradas modais e inferência com entradas modais específicas.

Experimentos abrangentes realizados no benchmark de preenchimento de profundidade KITTI demonstram que o método produz resultados robustos e supera outros métodos de última geração.

Dynamic Spatial Propagation Network for Depth Completion

Por fim temos o método que está atualmente em segundo lugar na posição do benchmark KITTI, sendo escolhido este e não o primeiro pelo fato da presença do artigo em link.

O preenchimento de profundidade guiado por imagem visa gerar mapas de profundidade densos com medições de profundidade esparsas e imagens RGB. Atualmente, as redes de propagação espacial (SPNs) são os métodos baseados em afinidade mais populares em preenchimento de profundidade, mas ainda sofrem com a limitação de representação da afinidade fixa e a suavização excessiva durante as iterações.

Nossa solução é estimar matrizes de afinidade independentes em cada iteração SPN, mas é um cálculo muito parametrizado e pesado. Este artigo apresenta um modelo eficiente que aprende a afinidade entre pixels vizinhos com uma abordagem dinâmica baseada em atenção. Ele desacopla a vizinhança em partes em relação a diferentes distâncias e gera recursivamente mapas de atenção independentes para refinar essas partes em matrizes de afinidade adaptativas.

Além disso, foi adotada uma operação de supressão de difusão (DS) para que o modelo convirja em um estágio inicial para evitar a suavização de profundidade densa. Finalmente, a fim de diminuir o custo computacional necessário, também introduzimos três variações que reduzem a quantidade de vizinhos e atenções necessárias, mantendo uma precisão semelhante.