

Disciplina: Residência em IA

Discente: Rodrigo Mendes de Carvalho - 202005505

Trabalho de Conclusão de Curso

Resumo dos artigos e dos tópicos levantados Computer Vision

1 - Camera Networks and Vision:

1.1 - A vision monitoring system for multipoint deflection of large-span bridge based on camera networking:

O artigo apresenta um sistema avançado de monitoramento por visão para a deflexão multiponto de pontes de grande extensão, denominado Camera Network System (CNS). Este sistema inovador foi desenvolvido para compensar os erros induzidos pelo movimento da câmera e permitir um monitoramento preciso e contínuo da deflexão de pontes em condições desafiadoras.

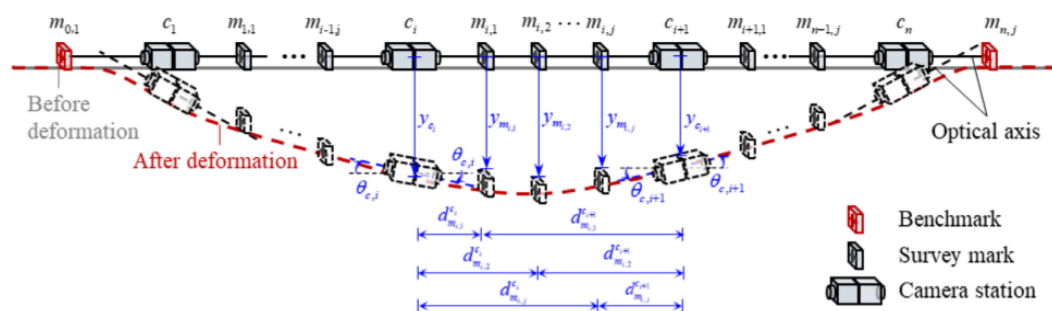


FIGURE 1 Schematic diagram of the proposed camera network system (CNS).

Introdução:

A necessidade de monitorar a deflexão de pontes de grande extensão de forma precisa e contínua é crucial para garantir a segurança e integridade dessas estruturas. O CNS proposto neste estudo visa superar as limitações dos métodos tradicionais de monitoramento e oferecer uma solução eficaz e inovadora para a engenharia civil.

Princípio de Medição e Componentes do CNS:

O CNS é baseado no princípio de videometria de deslocamento-relé, onde estações de câmeras duplas são posicionadas ao longo da ponte para capturar imagens de marcadores cooperativos. Cada estação de câmera é composta por dois módulos de câmera que atuam em conjunto, garantindo a precisão e consistência das medições. Além disso, um dispositivo de suplemento de iluminação infravermelha é integrado para permitir o monitoramento contínuo, mesmo em condições de baixa luminosidade.

Teste de Campo e Arranjo do CNS:

Um teste de campo foi realizado em uma ponte suspensa com 1038 metros de comprimento para validar a eficácia do CNS. As estações de câmeras foram instaladas dentro das vigas da ponte, permitindo a monitorização contínua da deflexão. Os resultados do teste demonstraram a capacidade do CNS em compensar os erros de movimento da câmera e fornecer medições precisas da deflexão da ponte em tempo real.



FIGURE 2 Camera network system (CNS) of bridge deflection monitoring.

Desafios e Discussões:

O estudo destaca desafios futuros, como o design otimizado do CNS para melhorar a precisão e robustez do sistema, a implementação em ambientes externos sujeitos a condições climáticas adversas e a análise pós-processamento dos dados coletados. Além disso, a integração com técnicas avançadas de aprendizado de máquina é apontada como uma área de pesquisa em desenvolvimento para aprimorar a análise dos dados de deflexão da ponte.

Conclusões e Resultados:

O CNS foi validado como uma solução viável e eficaz para o monitoramento da deflexão de pontes de grande extensão. O sistema demonstrou uma precisão submilimétrica nas medições e foi capaz de fornecer dados de deflexão comparáveis aos obtidos por sistemas tradicionais de monitoramento de longo prazo. Os resultados do teste de campo confirmaram a capacidade do CNS de compensar os erros de movimento da câmera e fornecer medições precisas da deflexão da ponte.

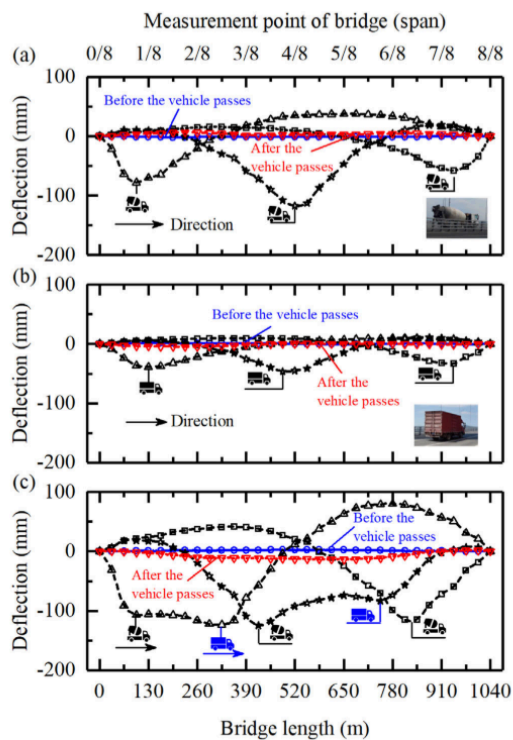


FIGURE 12 Effect of vehicle load on the full-span deflection evolution of bridge: (a) heavy truck, (b) medium truck, and (c) heavy and medium truck.

Considerações Finais e Agradecimentos:

O estudo foi apoiado por programas de pesquisa da China e destaca a importância do CNS para o monitoramento estrutural de pontes de grande extensão. O artigo conclui ressaltando a eficácia do CNS em compensar os erros de movimento da câmera e fornecer dados precisos e contínuos de deflexão, destacando seu potencial para aplicações futuras e desenvolvimentos na área de monitoramento estrutural.

Este resumo detalhado destaca a inovação, eficácia e potencial do Camera Network System proposto para o monitoramento de deflexão em pontes de grande extensão, evidenciando sua importância e contribuição para a engenharia civil e estrutural.

1.2 - Automated monitoring for security camera networks: promise from computer vision labs

Introdução:

A introdução destaca a importância crescente da monitorização automatizada para redes de câmeras de segurança, que visa melhorar a eficiência e a eficácia da vigilância por vídeo. Com o avanço da tecnologia de visão computacional, tornou-se possível desenvolver algoritmos sofisticados capazes de detectar eventos de interesse, identificar anomalias e resumir vídeos extensos de forma automatizada. Essa abordagem não apenas agiliza a revisão de imagens arquivadas, mas também melhora a capacidade de resposta a incidentes em tempo real.

Tarefa 1: Detecção de Anomalias

A detecção de anomalias desempenha um papel crucial na segurança, permitindo a identificação de comportamentos suspeitos que fogem do padrão normal. Algoritmos de visão computacional são treinados para reconhecer padrões incomuns e alertar os operadores de segurança sobre possíveis ameaças. Essa capacidade de identificar atividades não usuais em tempo real é essencial para prevenir incidentes e garantir a segurança de ambientes monitorados por câmeras de segurança.

Tarefa 2: Detecção de Eventos

A detecção de eventos envolve a identificação e classificação de ações específicas que são de interesse para a segurança, como agressões, furtos ou invasões. Por meio de algoritmos de CV, é possível treinar sistemas de vigilância para reconhecer automaticamente esses eventos e acionar alertas quando ocorrem. Essa capacidade de detecção de eventos em tempo real é fundamental para a tomada de decisões rápidas e eficazes em situações de emergência.

Tarefa 3: Sumários de Vídeo de Vigilância

A sumarização automática de vídeos é uma técnica essencial para condensar grandes volumes de imagens de vídeo em resumos concisos e informativos. Esses sumários destacam as atividades mais relevantes e eliminar informações redundantes, permitindo aos operadores de segurança revisar rapidamente horas de gravações em busca de eventos importantes. Algoritmos avançados de CV são empregados para criar esses sumários de forma eficiente e precisa, melhorando a eficácia da vigilância por vídeo.

Tarefa 4: Pesquisas Direcionadas por Consulta em Arquivos de Vídeo Longos

As pesquisas direcionadas por consulta permitem aos operadores de segurança buscar rapidamente segmentos de vídeo com características específicas, como a presença de indivíduos armados ou atividades suspeitas. Algoritmos de CV são aprimorados com informações fornecidas pelo usuário para melhorar a precisão das pesquisas e garantir a recuperação eficiente de dados relevantes em arquivos de vídeo extensos. Essa capacidade de pesquisa direcionada é essencial para investigações detalhadas e identificação de eventos críticos.

Conclusão

A monitorização automatizada para redes de câmeras de segurança, utilizando visão computacional, representa uma abordagem inovadora e eficaz para melhorar a segurança e a eficiência da vigilância por vídeo. A aplicação de algoritmos avançados de CV nas tarefas de detecção de anomalias, eventos, sumarização de vídeos e pesquisas direcionadas por consulta demonstra o potencial dessa tecnologia para aprimorar a segurança em ambientes monitorados. Com a contínua evolução da visão computacional, espera-se que essas técnicas se tornem ainda mais sofisticadas e eficientes, contribuindo para a proteção e o monitoramento eficaz de espaços públicos e privados.

2 - Sensors and Early Vision:

2.1- Current State of Hyperspectral Remote Sensing for Early Plant Disease Detection: A Review

Introdução:

A introdução discute a importância dos métodos de monitoramento remoto para resolver problemas relevantes na agricultura moderna, como o controle de pragas e doenças. Esses métodos têm potencial para fornecer informações úteis para práticas de manejo agrícola específicas. A introdução também menciona a existência de dois tipos de tecnologias de monitoramento remoto: passivas (como ópticas) e ativas (como LiDAR e radar). As tecnologias passivas ópticas são geralmente divididas em duas categorias com base na resolução espectral dos sensores utilizados: multiespectral e hiperespectral. A resolução hyperspectral mostra potencial como uma ferramenta não invasiva e não destrutiva para monitorar o estresse biológico e abiótico entre as tecnologias de monitoramento remoto passivo.

Essa técnica coleta e armazena informações do espectro de um objeto em um cubo de dados que contém informações espaciais e centenas de comprimentos de onda contíguos na terceira dimensão. A imagem hiperespectral com centenas de bandas espectrais pode fornecer retratos espectrais detalhados, permitindo a detecção de variações sutis no solo, nas copas ou em folhas individuais. Portanto, as imagens hiperespectrais podem ser usadas para resolver uma classe mais ampla de problemas para a determinação precisa e oportuna do estado fisiológico das culturas agrícolas.

A detecção precoce da propagação de doenças e surtos de pragas pode não apenas evitar perdas significativas de colheitas, mas também reduzir o uso de pesticidas e mitigar seus impactos negativos na saúde humana e no meio ambiente, melhorando assim o manejo integrado de pragas existentes.

Materials and Methods:

Esta seção discute o uso do sensoriamento remoto hiperespectral (HRS) na detecção de doenças em dendezeiros, o que é importante para o controle de pragas e doenças na indústria do óleo de dendê. A doença mais significativa é a podridão basal do caule (BSR) causada por *Ganoderma boninense*, que pode reduzir a

produtividade em 80%. Estudos demonstraram que o HRS pode discriminar entre dendezeiros saudáveis e infectados com BSR com alta precisão, usando várias técnicas, como PLS-DA, índices de vegetação e técnicas baseadas em red edge. A pesquisa também se concentrou no desenvolvimento de índices de vegetação espectral (SVIs) para detecção precoce de BSR em mudas de dendê.

Outra doença do dendezeiro discutida é a mancha laranja (OS), causada pelo viróide cadang-cadang do coco (CCCV), que matou milhões de coqueiros nas Filipinas. Embora atualmente seja de menor importância nos dendezeiros, estudos mostraram que o HRS pode detectar OS em dendezeiros sintomáticos usando vários SVIs.

A Tabela 1 resume os estudos sobre detecção precoce de doenças do dendê usando SHR, incluindo ano de publicação, cultura, tratamento, equipamentos utilizados, faixas estudadas, faixas importantes, tipo de estudo, referência e localização.

Table 1. Oil palm disease early detection by HRS.

Publication Year	Culture	Treat	Equipment	Studied Bands	Important Bands	Study Type	Reference	Location
2009	oil palm	basal stem rot	APOGEE spectroradiometer of unmentioned model	450–1100	715, 734, 791	field	[52]	Malaysia
2009	oil palm	basal stem rot	APOGEE spectroradiometer of unmentioned model	300–1000	462, 487, 610.5, 738, 749	field	[53]	Malaysia
2010	oil palm	basal stem rot	PP Systems Unispec-SC spectrometer	310–1130	670–715, 490–520, 730–770, 920–970	field	[50,51]	Indonesia
2011	oil palm	basal stem rot	APOGEE spectroradiometer of unmentioned model	350–1000	495, 495.5, 496, 651.5, 652, 652.5, 653, 653.5, 654, 654.5, 655, 655.5, 656, 656.5, 657, 657.5, 658, 658.5, 659, 659.5, 660, 660.5, 661, 908	field	[55]	Malaysia
2014	oil palm	basal stem rot	ASD spectrometer of unmentioned model	325–1040	not mentioned	field	[58]	Malaysia
2017	oil palm	basal stem rot	APOGEE spectroradiometer of unmentioned model	325–1000	495, 495.5, 496, 651.5, 652, 652.5, 653, 653.5, 654, 654.5, 655, 655.5, 656, 656.5, 657, 657.5, 658, 658.5, 659, 659.5, 660, 660.5, 661, 908	field	[56]	Malaysia
2017	oil palm	basal stem rot	GER 1500 spectrometer	273–1100	540–560, 650–780	field	[59]	Malaysia
2018	oil palm	basal stem rot	Specim spectrograph of unmentioned model	350–1000	650–750	field	[57]	Malaysia
2020	oil palm	basal stem rot	Cubert S185 camera	325–1075	800–950	greenhouse	[60]	Malaysia
2014	oil palm	orange spotting	ASD FieldSpec 4 spectrometer	300–1050	400–401, 404–405, 455–499, 500–599, 600–699, 700–712	field	[63,64]	Malaysia
2019	oil palm	orange spotting	ASD HandHeld 2 spectrometer	400–1050	601–630	field	[36]	Malaysia
2019	oil palm	orange spotting	ASD HandHeld 2 spectrometer	325–1075	680–780	field	[65,66]	Malaysia

Além disso, a seção menciona brevemente o uso do HRS na detecção de doenças dos citros, como o cancro bacteriano dos citros (CBC) e a doença do greening dos citros (HLB), que são ameaças significativas à indústria citrícola. No entanto, o foco da seção está nas doenças do dendezeiro.

Discussão:

No tópico de discussão do artigo, os autores destacam a importância de considerar os fatores abióticos na análise de estresse em plantas usando sensores hiperespectrais. Eles afirmam que não existe uma metodologia unificada para estudos hiperespectrais de doenças em plantas que leve em conta a influência desses fatores. Por esse motivo, sugerem a realização de experimentos em condições controladas de laboratório ou em estufas industriais para eliminar ou reduzir essas influências.

Além disso, os autores enfatizam a importância de indicar a cultura e o cultivo das plantas estudadas, bem como o patógeno utilizado para a inoculação. Eles também discutem a possibilidade de usar sensores hiperespectrais para detectar outras doenças em plantas além das já mencionadas no artigo, como a mancha laranja e a podridão da base do caule.

Outro ponto importante abordado na discussão é a influência da fluorescência da clorofila nos espectros de plantas e seus índices relacionados, que podem ser uma contribuição significativa para a solução do problema de detecção precoce de doenças em plantas. Os autores também mencionam a possibilidade de estudar a influência das características genóticas de um patógeno no perfil espectral de uma planta infectada.

Por fim, os autores discutem a importância dos efeitos das mudanças bioquímicas nos tecidos vegetais para a detecção precoce de doenças em plantas usando sensores passivos. Eles destacam que a refletância de luz em folhas de plantas é dependente de múltiplas interações biofísicas e bioquímicas, e que a detecção de elementos químicos individuais ou compostos químicos, incluindo compostos

voláteis, em plantas usando HRS pode ser crucial para a identificação precoce de doenças em plantas. No entanto, essa tarefa é difícil e pouco estudada.

Conclusão:

Na seção de conclusão do artigo, os autores reafirmam a possibilidade de detecção precoce de doenças em plantas usando hyperspectral remota sensoriamento (HRS). No entanto, eles também identificam algumas lacunas metodológicas e técnicas que precisam ser abordadas para garantir a replicabilidade dos experimentos.

Os autores sugerem que as diferenças nos espectros de plantas podem ser causadas por uma variedade de fatores abióticos e bióticos que causam estresse em plantas. Além disso, o fenótipo ou genótipo da planta hospedeira pode influenciar a manifestação da doença, dependendo do nível de resistência da planta. A presença de infecção mista também pode ser um fator importante que influencia o espectro da planta.

Em termos técnicos e físicos, os autores destacam a importância de considerar o modelo de propagação da luz solar durante experimentos de campo ou as características das fontes de luz artificial usadas em experimentos de laboratório, uma vez que o HRS é um método de sensoriamento remoto passivo que depende das condições da fonte de luz externa. Outros problemas técnicos podem incluir o uso incorreto do equipamento e a necessidade de calibrar adequadamente o sensor ou câmera hiperespectral para coleta de dados precisos.

Os autores concluem que a criação de bancos de dados de retratos hiperespectrais de plantas de diferentes culturas e cultivares expostas a diferentes patógenos deve ser adiada até que os princípios gerais de detecção remota de doenças em plantas usando HRS sejam desenvolvidos. Eles acreditam que a solução do problema da detecção precoce de doenças em plantas usando HRS requer a colaboração interdisciplinar de especialistas em fisiologia vegetal, fitopatologia, resistência vegetal, genômica, bioinformática, tecnologias da informação, análise de sistemas e óptica ou fotônica.

2.2 - RedEye: Analog ConvNet Image Sensor Architecture for Continuous Mobile Vision

Introdução:

A recente emergência de dispositivos wearables trouxe a ideia de "mostrar aos computadores o que você vê", ou visão contínua móvel. No entanto, a barreira para a eficiência energética é desafiadora. Tarefas de visão contínua esgotam a bateria do Google Glass em 40 minutos. A melhoria da eficiência energética dos circuitos digitais por meio de tecnologia de processo e otimização de nível de sistema pode continuar, mas medições recentes apontam para um gargalo fundamental na eficiência energética: o sensor de imagem, especialmente sua circuiteria de leitura analógica.

Esse gargalo analógico é fundamental por duas razões. Em primeiro lugar, em comparação com os circuitos digitais, a eficiência energética da leitura analógica melhora muito mais lentamente ao longo das gerações tecnológicas. Em segundo lugar, enquanto a tendência na comunidade de circuitos de estado sólido é mover funcionalidades do domínio analógico para o digital, a leitura analógica será necessariamente existente como uma ponte do mundo analógico físico para o reino digital.

Para abordar esse gargalo, nossa ideia-chave é empurrar a processamento antecipado no domínio analógico para reduzir a carga de trabalho da leitura analógica, resultando em nosso design do RedEye, um sensor de imagem convolutivo analógico para visão móvel contínua. RedEye descarta dados crus, exportando recursos gerados por processamento de uma Rede Convolutacional (ConvNet) no domínio analógico. Isso economiza energia não apenas na leitura analógica do sensor, mas também em todo o sistema.

O design do RedEye aborda os seguintes desafios importantes: (i) os circuitos analógicos sofrem de alta complexidade de design, desencorajando a portabilidade entre nós de tecnologia de processamento. RedEye restringe a complexidade com um design modular que facilita a reutilização física dentro de um design de circuito e reutilização algorítmica por meio de fluxo de dados cíclico. (ii) A processamento analógica está sujeita a acúmulo de ruído, o que afeta a precisão da tarefa. RedEye identifica os principais compromissos entre ruído-energia para sacrificar energia para amenizar o ruído durante o processamento. RedEye fornece mecanismos para ajustar parâmetros de ruído para eficiência e precisão em tempo de execução. (iii) Os desenvolvedores que usam o RedEye requerem uma estimativa da precisão da tarefa da ConvNet e do consumo de energia sob a influência do ruído. Fornecemos

um framework de simulação para ajudar os desenvolvedores a ajustar o equilíbrio de precisão e eficiência ao executar ConvNets no RedEye.

Background:

A. Image sensing

Dispositivos móveis modernos usam sensores de imagem CMOS para fotografia e videografia. Embora mais eficientes que CCD, eles ainda consomem centenas de miliwatts devido ao seu analógico oneroso leitura [4].

Os sensores de imagem modernos consistem em três componentes principais: matriz de pixels, leitura analógica e lógica digital. A matriz de pixels converte a luz em elétrons e depois em tensões, consumindo pouca energia [5]. A leitura analógica amplifica sinais analógicos da matriz de pixels e os converte em valores digitais. A lógica digital mantém a operação do sensor, gerenciando a geração de tempo, scanner e motoristas de quadro e interface de dados. Os sensores modernos empregam uma arquitetura de coluna para leitura analógica; pixels em cada coluna compartilham circuitaria dedicada. Amplificadores de coluna amostram e amplificam pixels crus para aproveitar o alcance de sinal disponível e superar o ruído; eles consomem energia estática para polarizar os transistores operando linearmente. Conversores A/D de coluna digitalizam sinais amplificados, consumindo energia estática e dinâmica. A leitura analógica consome 50%-75% da energia total em designs de sensores recentes [5]-[7]. Isso porque (i) todo o sinal de imagem deve passar por leitura analógica e (ii) leitura analógica é projetada para alta fidelidade.

Destacamos que a otimização digital, como hardware de baixa potência para redução de dados [8], descarte antecipado ou computação offloading, não afeta a energia consumida pela leitura analógica. Isso é o fundamento da nossa proposta para mover a processamento antes leitura analógica.

B. Eficiência de computação analógica

A computação analógica é mais eficiente do que a computação digital para operações básicas, como adição e multiplicação, com a energia por operação de implementações analógicas sendo ordens de magnitude menor [9]. Ao contrário da computação digital, que usa várias cargas binárias (bits) para representar o estado do circuito, a computação analógica usa uma única carga para representar um valor. Isso reduz a contagem de hardware; adição analógica só precisa de um interconexão para juntar duas correntes, enquanto um adder digital de 16 bits requer mais de 400 transistores.

Mais importante, trocas lucrativas de energia, eficiência e fidelidade de sinal podem ser feitas no domínio analógico, relevantes para tarefas de visão resistentes ao ruído. Isso porque o mantendo estado através da capacitância C é suscetível ao ruído térmico $V_{2n} = \sqrt{kT/C}$ [12]. Embora aumentar C desencoraje o ruído, a energia necessária para encher a célula de memória aumenta proporcionalmente. Portanto, a escolha de C é um compromisso fundamental entre a fidelidade e a eficiência da computação analógica. Aqui nós identificamos os compromissos de energia-ruído inerentes à computação analógica.

Memória: Como um pipeline analógico deve ser construído em estágios para facilitar a configurabilidade e manutenção, a memória analógica é indispensável para buffers inter-estágio. As células de memória usam capacitores para manter estados e, portanto, exibem compromissos de energia-ruído na leitura e gravação de valores.

Aritmética: A aritmética em circuitos de carga, como em RedEye, geralmente requer um amplificador operacional (op amp), que induz ruído térmico $\propto 1/C$ e energia $\propto C$.

C. ConvNets

Convolutional Neural Networks (ConvNets) são usados para tarefas de visão contínua em RedEye devido à sua eficácia em detecção de objetos, reconhecimento de escrita manual e detecção facial [18]. Essa proeminência resultou em pesquisa acadêmica e industrial ativa na implementação de ConvNets para tarefas de visão.

ConvNets opera processando dados tridimensionais através de camadas sequenciais de processamento para gerar recursos de visão. Cada camada consiste em um conjunto de neurônios; cada neurônio recebe múltiplos valores de entrada de um campo receptivo local de neurônios da camada anterior. O neurônio opera nestes valores para gerar uma única saída. Camadas de neurônios vêm em uma pequena variedade de tipos. Aqui cobrimos os principais tipos de camadas:

Neurônios de camada convolucional multiplicam um campo receptivo tridimensional de entradas com um kernel de pesos para gerar uma saída. Neurônios em uma camada convolucional compartilham pesos de kernel para executar seu processamento. Conjuntos múltiplos de pesos por camada geram canais de profundidade de dados. Um processo de retropropagação treina pesos para alterar a operação de convolução.

Neurônios da camada de não linearidade usam funções de ativação para introduzir saturação não linear dos outputs de camadas convolucionais. Funções sigmoid, tangente hiperbólica e funções de reta foram usadas para criar não linearidade.

Neurônios da camada de pooling máximo recebem recursos gerados de um campo receptivo de saída processada e identificam o recurso com a resposta máxima. Isso permite que a rede seja robusta a pequenas mudanças no espaço de recursos.

Essas camadas, além de outras camadas, como camadas de normalização para restringir a amplitude do sinal, podem ser encadeadas em ordem sequencial variada, invocando cada camada várias vezes para criar estruturas profundas. Por exemplo, a rede de reconhecimento de imagem ConvNet AlexNet usa 6 camadas convolucionais, 7 camadas de não linearidade, 3 camadas de pooling, 2 camadas de normalização. AlexNet também inclui 7 "outras" camadas no final de sua operação.

Devido à generalização de Redes treinadas ConvNets em entradas de imagem, eles são naturalmente robustos ao ruído [20]. Nós estudamos e exploramos a robustez para sfruttare os compromissos de energia-ruído em RedEye.

Redeye System Architecture (III. REDEYE SYSTEM ARCHITECTURE)

O terceiro tópico descreve a arquitetura do sistema redeye. O sistema redeye consiste em um coprocessador com processadores específicos e um sistema de memória altamente programável (HPMs). O coprocessador redeye é projetado para executar operações de CNN específicas de maneira eficiente. Ele contém processadores específicos, como um processador de ponto fixo (PFP) para processamento de matrizes, e um processador de ponto flutuante (FPP) para processamento de pontos flutuantes. Além disso, o sistema redeye utiliza um sistema de memória altamente programável (HPMs) para armazenar dados de imagem e pesos de rede.

Redeye Circuit Design (IV. REDEYE CIRCUIT DESIGN)

O quarto tópico aborda o design do circuito redeye. O design do circuito redeye é projetado para ser compacto, energeticamente eficiente e com alto desempenho. Ele utiliza um design de soma de produtos de múltiplos portadores (MCCA) para o processamento de matrizes. O design do circuito também utiliza uma topologia de árvore de pesquisa de intervalos para acelerar o processamento de pontos flutuantes. Além disso, o circuito redeye utiliza técnicas de baixa potência, como

gerenciamento de tensão de alta e baixa velocidade e multiplexação de alimentação.

Evaluation (V. EVALUATION)

O quinto tópico avalia o desempenho do sistema redeye em comparação com a solução baseada em FPGA (Field Programmable Gate Array) existente. O desempenho do sistema redeye é avaliado em termos de área de silício, consumo de energia, desempenho e eficiência. Os resultados mostram que o sistema redeye oferece um desempenho melhor em relação à solução baseada em FPGA em termos de desempenho e eficiência.

Related Work (VI. RELATED WORK)

O sexto tópico apresenta as técnicas e soluções relacionadas à otimização do processamento de imagens em dispositivos portáteis. Ele menciona algoritmos de comutação rápida de alta velocidade, como o FPGA (Field Programmable Gate Array) e o ASIC (Application-Specific Integrated Circuit), e arquiteturas de baixa potência, como o HPMS (Highly Programmable Memories). Além disso, o tópico menciona técnicas de otimização de software, como a otimização de código e a otimização de memória.

Concluding Remarks (VII. CONCLUDING REMARKS)

O sétimo tópico resume os principais resultados do artigo e apresenta as conclusões. O artigo propõe uma abordagem para o desenvolvimento de um coprocessador chamado "redeye" para acelerar o processamento de imagens em CNNs. O sistema redeye é projetado para ser compacto, energeticamente eficiente e com alto desempenho. Os resultados mostram que o sistema redeye oferece um desempenho melhor em relação à solução baseada em FPGA em termos de desempenho e eficiência. Além disso, o artigo menciona as técnicas e soluções relacionadas à otimização do processamento de imagens em dispositivos portáteis.

3 - Machine Learning Technologies for Vision:

3.1 - Machine Learning in Computer Vision Asharul Islam Khan a*, Salim Al-Habsib

Introdução:

A introdução destaca a crescente importância do aprendizado de máquina e da visão computacional, visando capacitar os computadores com habilidades humanas de detecção e compreensão de dados. Essas tecnologias têm aplicabilidades variadas, desde a Internet das Coisas até interfaces cerebrais humanas. Métodos como aprendizado supervisionado, não supervisionado e semi-supervisionado são fundamentais, utilizando algoritmos como máquinas de vetores de suporte e KNN. Empresas oferecem soluções de reconhecimento de fala, análise de texto e classificação de imagens, acessíveis através de APIs. A detecção e análise de objetos são cruciais para diversas aplicações, como evitar colisões de trânsito e reconhecimento emocional.

Tecnologias como Tensor Flow e OpenPose são utilizadas para detecção de objetos. O aprendizado supervisionado de redes neurais convolucionais é destacado, juntamente com o desafio da anotação de dados. O serviço de aprendizado de máquina em nuvem é uma tendência, com empresas como Amazon, Microsoft e Google oferecendo plataformas para tal. O objetivo da pesquisa é analisar aplicativos de aprendizado de máquina em visão computacional, com base em uma revisão bibliográfica que resultou em 20 artigos selecionados para este estudo. O trabalho é dividido em cinco seções, abrangendo estudo de base, categorização de aplicativos existentes, resultados, discussões e conclusões.

Background study

A visão computacional e o aprendizado de máquina são duas áreas importantes de pesquisa recente. A visão computacional utiliza mapeamento de imagem e padrões para encontrar soluções, considerando uma imagem como uma matriz de pixels. Automatiza tarefas de monitoramento, inspeção e vigilância. O aprendizado de máquina, um subconjunto da inteligência artificial, resulta na análise/rotulagem automática de vídeos. Existem três abordagens para aprendizado de máquina e visão computacional: supervisionado, não supervisionado e semi-supervisionado. O aprendizado supervisionado requer dados de treinamento rotulados, enquanto o semi-supervisionado possui alguns dados rotulados e outros não.

Problemas do mundo real muitas vezes se encaixam na categoria de aprendizado não supervisionado, onde padrões evoluem com base em agrupamento. Os paradigmas de aprendizado de máquina para visão computacional incluem máquinas de vetores de suporte, redes neurais e modelos gráficos probabilísticos. As Redes Neurais Convolucionais (CNNs) têm ganhado popularidade devido a conjuntos de dados acessíveis, GPUs e técnicas de regularização. A biblioteca OpenCV é integrada com diversas linguagens de programação e plataformas para processamento de imagem, análise de vídeo, detecção de objetos e aprendizado de máquina.

Aprendizado de Máquina em Visão Computacional:

O estudo explorou diversas aplicações do aprendizado de máquina na visão computacional, abrangendo áreas como segmentação, extração de características, refinamento de modelos visuais, correspondência de padrões, representação de formas, reconstrução de superfícies e modelagem para ciências biológicas.

Algumas das aplicações incluem interpretação de dados em imagens de detecção de carros e pedestres, classificação automática de falhas em trilhos ferroviários, diferenciação de variedades de manga com base em atributos de tamanho, extração de informações gráficas e textuais de imagens de documentos, reconhecimento de gestos e rostos, visão de máquina, reconhecimento de caracteres e dígitos manuscritos, sistemas avançados de assistência ao motorista, estudos comportamentais, e estimativa de cinemática corporal completa humana para um ciclista e estimativa de pose.

Além disso, o estudo aborda aplicações em medicina, como detecção de hemorroidas, detecção de sangramento, aprimoramento de imagens endoscópicas e suporte à decisão clínica. Outras áreas de aplicação incluem ciências biológicas, meteorologia, detecção e contagem de tráfego, análise de desempenho em esportes, avaliação do estado de ferramentas de máquinas, manutenção preditiva e muito mais. As aplicações são detalhadas e agrupadas em uma tabela para facilitar a compreensão e o acesso.

Resultados e Discussão:

Os resultados e discussões deste estudo destacam a crescente importância do aprendizado de máquina e da visão computacional em diversas áreas. As técnicas de aprendizado de máquina e visão computacional reduziram custos, esforços e tempo em engenharia, ciência e tecnologia. Um sistema automatizado baseado em

aprendizado de máquina e visão computacional detecta as emoções humanas, prevê atividades humanas por meio de rotulagem e reconhecimento de padrões, e analisa o desempenho de equipes e jogadores individuais em esportes profissionais. Além disso, essas técnicas são utilizadas em indústrias para manutenção preditiva, o que impacta significativamente a eficácia e eficiência das unidades de fabricação. A visão computacional e o aprendizado de máquina aplicados em dados de câmeras públicas e dispositivos inteligentes ajudam na previsão e monitoramento do tráfego nas cidades. Os resultados mostram que as áreas de pesquisa avançadas incluem ciências biológicas, atividade humana, gestão de tráfego e esportes profissionais. A evolução do aprendizado de máquina, desde métodos tradicionais de reconhecimento de padrões até técnicas avançadas de compreensão de imagens, contribui para a dinâmica em mudança dos sistemas de visão computacional. Embora a visão computacional possa interpretar e extrair informações de áudio e vídeo de forma independente, o aprendizado de máquina adiciona a capacidade preditiva dos dados já processados. As aplicações de aprendizado de máquina em visão computacional têm saídas variadas dependendo do domínio, e os principais temas de pesquisa incluem classificação de objetos, detecção de objetos e reconhecimento de sequências. No entanto, a detecção e rastreamento de objetos ainda representam um desafio aberto na visão computacional. A qualidade da saída do aprendizado de máquina depende da precisão preditiva, recall e precisão.

Conclusão:

O estudo conclui que a pesquisa comercial e acadêmica em visão computacional está crescendo, e a síntese entre aprendizado de máquina e visão computacional tem sido fundamental para entender problemas complexos. O futuro trabalho incluirá a avaliação da precisão dos algoritmos de aprendizado de máquina em visão computacional.

3.2 - Deep learning for consumer devices and services: Pushing the limits for machine learning, artificial intelligence, and computer vision

Lemley, Joseph; Bazrafkan, Shabab; Corcoran, Peter

Introdução:

O texto introduz o conceito de Deep Learning, destacando seu rápido crescimento devido à disponibilidade de grandes conjuntos de dados e hardware especializado. Ele destaca a aplicabilidade do Deep Learning em uma ampla gama de problemas,

incluindo a melhoria de dispositivos e serviços de consumo. Além disso, o texto explica os fundamentos das redes neurais artificiais (ANN), incluindo sua capacidade de aprendizado e generalização.

Ele discute o treinamento de redes neurais, abordando conceitos como underfitting e overfitting, além de métodos de avaliação de modelos. O texto também diferencia entre aprendizado supervisionado e não supervisionado e explora a estrutura interna das redes neurais, descrevendo camadas, neurônios e mecanismos de conexão. Finalmente, ele destaca o recente renascimento das redes neurais devido a avanços tecnológicos e ao crescimento do Big Data.

Redes Neurais Convolucionais

As Redes Neurais Convolucionais (CNN) processam sinais analógicos, como imagens e sons, convertendo-os para o formato digital através da conversão analógico-digital (A2D). Elas operam em espaços dimensionais altos, representando amostras em um espaço de características. As CNNs utilizam camadas de convolução, que aprendem filtros durante o treinamento para extrair características importantes dos dados. Essas camadas aplicam o operador de convolução, que compara o sinal de entrada com um filtro para detectar padrões.

As CNNs também incluem camadas de pooling, que reduzem a dimensionalidade dos dados e fornecem invariância a transições. Além disso, as CNNs empregam funções de ativação não-lineares para modelar sistemas não-lineares. As arquiteturas de CNNs podem incluir camadas densas completamente conectadas após as camadas de convolução, formando redes neurais profundas genéricas. Outras arquiteturas de destaque são as Redes Neurais Totalmente Convolucionais (FCN), que não possuem camadas densas, os Autoencoders, que aprendem representações comprimidas dos dados, e as Redes Neurais Recorrentes (RNN), projetadas para dados sequenciais.

Estado da Arte Hoje:

As tecnologias habilitadoras principais estão impulsionando o avanço das redes neurais profundas e suas aplicações. As GPUs (unidades de processamento gráfico) desempenham um papel crucial no treinamento dessas redes, sendo significativamente mais rápidas que as CPUs para tarefas altamente paralelizáveis, como operações matriciais. Enquanto as CPUs são otimizadas para problemas sequenciais, as GPUs são projetadas para processamento gráfico paralelo, o que as torna ideais para deep learning. Com centenas ou milhares de núcleos, as GPUs treinam redes neurais profundas em uma magnitude de ordem mais rápida do que

as CPUs. A série Pascal da NVIDIA representa uma melhoria significativa no desempenho e no consumo de energia em comparação com modelos anteriores.

As GPUs podem ser combinadas para treinamento mais rápido, mas a escolha do dispositivo depende da memória e do número de núcleos necessários para a tarefa. Além das GPUs, surgem dispositivos de hardware personalizados para deep learning, como o "fathom neural compute stick" da Movidius, que consome significativamente menos energia que uma GPU equivalente. Este tipo de dispositivo é ideal para implementações em objetos inteligentes, como robôs e drones.

O software também desempenha um papel fundamental no desenvolvimento e na implementação de modelos de deep learning. Frameworks como Caffe, Theano e TensorFlow simplificam o design e a implantação de redes neurais profundas, oferecendo abstrações de alto nível e facilitando o desenvolvimento. O TensorFlow, desenvolvido pelo Google, está crescendo rapidamente em popularidade devido à sua facilidade de uso em múltiplas GPUs.

Essas tecnologias permitiram uma ampla gama de aplicações práticas, desde tradução visual instantânea até reconhecimento facial, passando por reconhecimento de voz e transformação de fotos em obras de arte. Empresas como Google, Nuance e Apple estão incorporando deep learning em seus produtos para melhorar a experiência do usuário e oferecer novos recursos, como reconhecimento de voz e imagem em tempo real. Além disso, frameworks como Keras tornam o desenvolvimento acessível mesmo para iniciantes, permitindo que uma ampla variedade de problemas seja abordada por meio de deep learning.

Em resumo, o estado atual da arte em deep learning é caracterizado por uma combinação de hardware avançado, software de alto nível e uma ampla gama de aplicações práticas, impulsionando o campo em direção a novos avanços e inovações em uma variedade de domínios.

Conclusão:

Hoje, a aprendizagem profunda está sendo usada em nossos celulares, em nossos carros, em tablets e computadores. Ela ampliou os limites do que é possível para tarefas como segmentação de imagem, detecção de objetos, reconhecimento facial, análise de voz, detecção de emoções e reconhecimento de gênero.

Por que a aprendizagem profunda de repente catalisou pesquisas em tantos campos? Bem, é uma combinação de muitos fatores: o surgimento recente de

hardware computacional baseado em GPU altamente acessível e denso forneceu os meios para processar conjuntos de dados muito grandes e implementar as metodologias de treinamento avançadas necessárias para desenvolver CNNs precisas; a disponibilidade generalizada de GPUs em dispositivos de hoje, combinada com serviços de processamento de dados baseados em nuvem, fornece os meios para aplicar essas arquiteturas de CNN a aplicativos cotidianos, como processamento de voz ou imagem. Grandes conjuntos de dados fornecem o combustível para impulsionar a atividade de pesquisa e refinar resultados até o ponto em que as soluções de aprendizagem profunda geralmente superam até mesmo as melhores ferramentas de reconhecimento de padrões projetadas pelo homem.

Hoje, estamos em um ponto em que muitos novos problemas podem ser abordados por meio de técnicas de aprendizagem profunda - muitos desses são problemas antigos, como reconhecimento de voz, que nunca foram bons o suficiente para sair do laboratório e entrar no uso diário. Mas este ano vimos vários lançamentos de 'alto-falantes inteligentes' que podem controlar sua casa - e por trás desses novos dispositivos estão uma série de tecnologias de aprendizagem profunda - analisando suas necessidades, traduzindo suas solicitações de voz e coordenando a logística necessária para entregar tudo em sua porta ou na sua TV.

Para começar a resolver seu próprio problema, você precisará de uma GPU de ponta - a mesma tecnologia usada nos últimos PCs de jogos - e a maioria do software principal está disponível gratuitamente na Internet. Existem vários pacotes de software em tendência no campo da aprendizagem profunda, incluindo, mas não se limitando a Theano (em Python), Lasagne (em Theano), Tensor Flow (em Python e C++), Caffe (em Python e MATLAB) e MatConvNet (em MATLAB).

4 - Object Recognition

4.1 - Research and Application of Visual Object Recognition System Based on Deep Learning and Neural Morphological Computation Le Yang^{1, *}, Han Wang², Jiajian Zheng³, Xin Duan⁴, Qishuo Cheng⁵

Introdução:

A introdução destaca os desafios enfrentados pelos sistemas de computação tradicionais, baseados em arquiteturas de Von Neumann, como limitações na

velocidade de processamento e aumento dramático no consumo de energia devido à separação entre unidades de processamento e armazenamento. Os pesquisadores propõem explorar a computação neuromórfica baseada em sinapses artificiais como uma solução promissora para esses desafios. Essa abordagem oferece vantagens significativas, como alta velocidade de processamento e baixo consumo de energia. O foco do artigo está na aplicação dessas tecnologias no reconhecimento visual de objetos, destacando os avanços recentes em visão computacional e deep learning.

Trabalhos Relacionados:

Esta seção explora o conceito de computação neuromórfica e sua aplicação em redes neurais profundas (DNNs) para reconhecimento de padrões em grandes conjuntos de dados. Os autores discutem o desenvolvimento de hardware neuromórfico, incluindo memristores e sistemas VLSI, e como essas tecnologias estão sendo combinadas com redes neurais profundas para melhorar a eficiência de tarefas como reconhecimento de imagens e classificação. São abordadas as limitações históricas da computação neuromórfica, bem como os avanços recentes que a tornam uma opção mais viável.

Metodologia:

O artigo descreve detalhadamente a metodologia utilizada para investigar o comportamento resistivo digital e analógico de folhas de WSe₂ em memristores, juntamente com a simulação de funções sinápticas básicas e computação neuromórfica. Métodos de caracterização, como microscopia eletrônica de varredura (SEM), espectroscopia Raman e espectroscopia de fotoelétrons de raios-X (XPS), são empregados para analisar as propriedades físicas das folhas de WSe₂. O processo de crescimento de nano lâminas de WSe₂ é detalhado, assim como os resultados da caracterização por diferentes técnicas.

Conclusão:

Os autores concluem destacando as perspectivas promissoras da integração entre computação neuromórfica e deep learning em sistemas de reconhecimento visual de objetos. Eles enfatizam como essa integração pode melhorar a eficiência e a precisão dos sistemas de reconhecimento visual, acelerando o progresso da tecnologia de visão computacional. Além disso, apontam que essa integração estabelece uma base sólida para o desenvolvimento de sistemas inteligentes de percepção e cognição. Os autores sugerem que futuras pesquisas devem explorar

ainda mais a interseção entre computação neuromórfica e deep learning para avançar ainda mais a área de reconhecimento visual de objetos.