ObsevartóR!O2016 - interseções entre arte e técnicas de Deep Learning

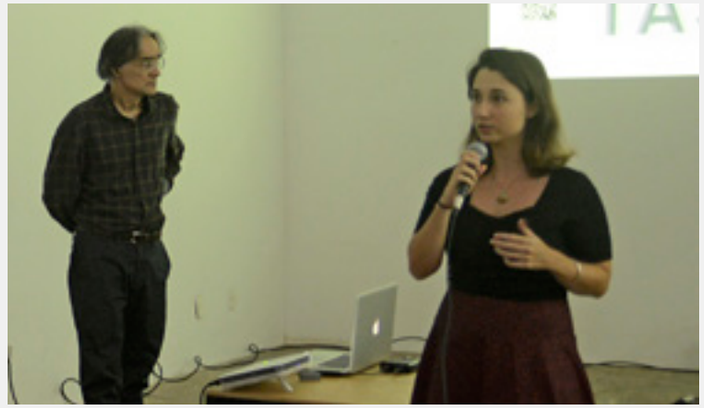
Julia Giannella, Luiz Velho, Juliano Kestenberg, Vitor Guerra e Djalma Lucio

Resumo: A produção e circulação de grandes volumes de imagens digitais nas redes sociais inauguram novas abordagens de pesquisa para disciplinas interessadas na prática criativa humana. Este trabalho revisitou o acervo de 180 mil imagens sobre a Rio-2016 compartilhadas no Twitter e coletadas para o projeto OBSERVATÓR!O2016 (http://oo.impa.br), do Laboratório VISGRAF, e explorou formas automatizadas de reagrupar as imagens e remixá-las em novos produtos audiovisuais a partir do reconhecimento de objetos, cenas e faces em fotografias possibilitado por técnicas de deep learning (aprendizagem profunda). Para o Encontro Indisciplinas, apresentamos dois vídeos simultâneos no formato de projeções. A experiência audiovisual das projeções pretende gerar uma reflexão sobre as implicações do uso de máquinas inteligentes no exercício artístico.

Palavras-chave: máquinas inteligentes; deep learning; imagem; midiarte; rio-2016.

www.visgraf.impa.br/dl_rio2016

https://youtu.be/kDDcKEq6U1s



Julia Giannella é graduada em Desenho Industrial pela UFRJ, mestre em Ciências da Comunicação pela USP e doutoranda em Design pelo PPDESDI–UERJ. É assistente de pesquisa no VISGRAF-IMPA.

Luiz Velho é professor e pesquisador titular do IMPA e líder do VISGRAF. Mestre em Computação Gráfica pelo MIT Media Lab e doutor em Ciências da Computação pela Universidade de Toronto.

Juliano Kestenberg é graduado em Desenho Industrial pela ESDI–UERJ, mestre em Design pelo PPDESDI–UERJ e doutorando em Artes Visuais no PPGAV–UFRJ. É assistente de pesquisa no VISGRAF-IMPA.

Vitor Guerra é graduado em Processamento de Dados, mestre em Sistemas e Computação pelo IME e doutor em Engenharia Informática pela Universidade de Coimbra, Portugal. É pós-doc no VISGRAF-IMPA.

Djalma Lucio é graduado em Ciência da Computação pelo Centro Universitário Senac-SP, mestre em Informática pela PUC-RIO. É desenvolvedor e sysadmin no VISGRAF-IMPA.

Introdução

A produção e circulação de grandes volumes de imagens digitais nas redes sociais inauguram novas abordagens de pesquisa para disciplinas interessadas na prática criativa humana. Se, por um lado, o acesso a esse vasto repertório imagético é facilitado pelos avanços em tecnologias de coleta e armazenamento de dados visuais, por outro lado, o acesso em si não é suficiente para propor problemas de análise e síntese de imagens. Volumosas coleções de artefatos visuais justificam a urgência de se explorar mecanismos inteligentes capazes de identificar e agrupar imagens não apenas a partir de seus metadados textuais, mas principalmente a partir de metodologias de análise e processamento de dados visuais.

O Laboratório de Visão e Computação Gráfica (VISGRAF) – pertencente ao Instituto Nacional de Matemática Aplicada (IMPA) – desenvolve, dentre outras atividades, pesquisa e projetos em visão computacional, área da computação que lida com maneiras de analisar e sintetizar imagens a partir do reconhecimento de formas, estilos e propriedades visuais em imagens e vídeos. Nesse sentido, o Laboratório identificou a oportunidade de explorar um acervo de aproximadamente 180 mil imagens coletadas do Twitter para o projeto OBSERVATÓR!O2016,[1] interessado em compreender as múltiplas perspectivas sobre as Olimpíadas Rio-2016 compartilhadas através de textos e imagens na rede social.

A partir da identificação desta oportunidade de pesquisa, o projeto original se desdobrou em uma investigação exploratória que resultou no site "OBSERVATÓR!O2016: experiências em **deep learning**".[2] No site, três problemas tradicionais do campo da visão computacional são explorados a partir de procedimentos de aprendizagem profunda (**deep learning**), uma técnica de implementação de aprendizagem por máquina (**machine learning**):

- Classificação de imagens e reconhecimento de objetos em fotografias;
- Transferência de estilo artístico;
- Geração automática de vídeo.

Os três problemas acima elencados foram explorados a partir da coleção de imagens da Rio-2016[3] e resultaram em três experiências apresentadas no site, a saber:

- **Mosaico da tocha**: coleção de imagens do tour da tocha olímpica visualizada na forma de um mosaico interativo;
- **Globo de vídeos**: **slideshows** musicais que agrupam imagens por esportes olímpicos e atletas;
- **Impressões sonoras**: declarações de atletas e condutores da tocha combinados com áudios que marcaram a Rio-2016 e imagens com aplicação de estilos artísticos.

Mais informações sobre o procedimento criativo-

1 GIANNELLA, J. R.; VELHO, L. "OBSERVATÓR!O2016". **Technical Report TR-08-2016**, Laboratório VISGRAF - IMPA, 2016. Disponível em: http://www.visgraf.impa.br/Data/RefBib/PS_PDF/tr-08-2016/tr-08-2016.pdf.

2 O site está disponível em www.visgraf.impa.br/dl_rio2016.

3 A coleção de imagens pode ser acessada por dia e categoria em http://oo.impa.br/dimagens/. As imagens também estão armazenadas em um banco de dados do VISGRAF. Para acesso ao banco de dados através de um API Rest, entre em contato com lvelho@impa.br.

metodológico que resultou nas experiências podem ser encontradas na seção **Aspectos Metodológicos** do site .[4] A seguir, descrevemos o processo de adaptação das experiências em **deep learning** para apresentação na forma de vídeo-projeções no Encontro **Indisciplinas**, realizado na Casa França-Brasil entre os dias 22 e 25 de novembro de 2016.

Desenvolvimento das vídeo-projeções

Em razão da convocatória de trabalhos para o **Indisciplinas** – 4º Encontro dos Programas de Pós-Graduação em Artes Visuais do Estado do Rio de Janeiro –, submetemos as experiências envolvendo **deep learning** e as imagens da Rio-2016 para apresentação no formato de uma exibição artística. A proposta foi aceita e propusemos a exposição de duas vídeo-projeções que mesclam as três experiências citadas na seção anterior.

Para a exibição artística, reunimos os **slideshows** musicais que compõem a experiência **Globo de Vídeos** em dois únicos vídeos para serem projetados simultaneamente em paredes distintas (veja seção **Preparação dos vídeos**). Os vídeos receberam uma nova trilha sonora, editada a partir de novos áudios e músicas selecionadas (veja seção **Trilha sonora**).

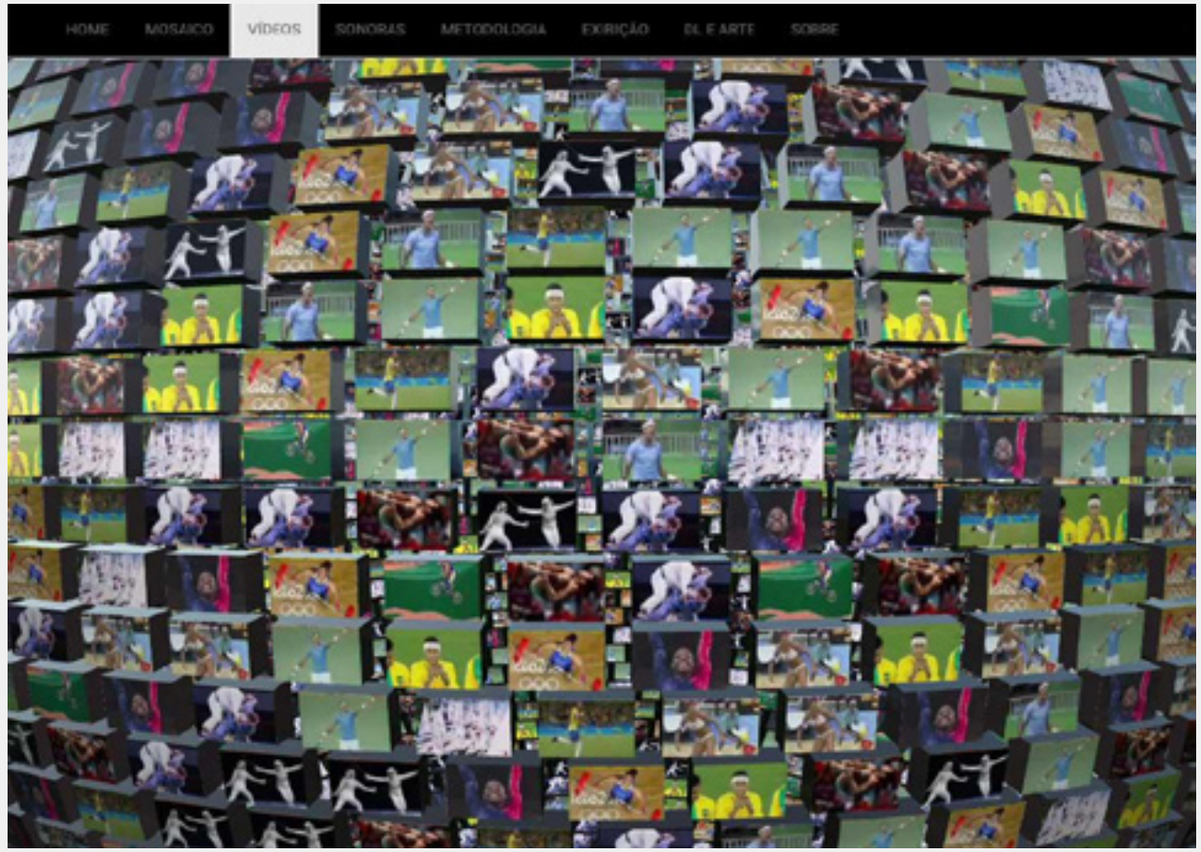


Figura **1 Globo de Vídeos**, no ambiente web. Fonte: reprodução.

Preparação dos vídeos

A experiência **Globo de Vídeos** (Figura 1) apresenta, no ambiente web, uma visualização em formato de esfera 3D contendo a miniatura de dezenas de vídeos. A esfera pode ser manipulada com

4 A seção **Aspectos Metodológicos** está disponível em http://lvelho.impa.br/dl_rio2016/metodologia.html.

interações de **pan** e **zoom**, e cada miniatura, ao ser clicada, inicializa um **slideshow** musical. Ao todo, o **Globo de Vídeos** possui doze **slideshows** distintos, caracterizados pelas imagens de esportes olímpicos ou atletas. São eles: **Beach Volley**, **Cycling**, **Fencing**, **Football**, **Juan Martín Del Potro**, **Judo**, **Medals**, **Olympic Gymnastics**, **Sailing**, **Simone Biles**, **Tennis** e **Weightlifting**.

A preparação dos **slideshows** envolveu dois problemas da área da visão computacional, considerados na seguinte ordem: 1) classificação de imagens e 2) geração automática de vídeo. Para classificar as imagens automaticamente de acordo com as modalidades olímpicas ou premiação de medalhas nelas retratadas, implementamos uma técnica de retreinamento supervisionado a partir de uma rede neural chamada **Inception-v3**.[5]

Já para a geração automática de vídeo – a síntese dos **slideshows** propriamente dita –, recorremos a recursos automatizados oferecidos pela empresa Apple. O novo sistema operacional da Apple (iOS 10), destinado a seus dispositivos móveis, traz uma forte presença de inteligência artificial em seus aplicativos nativos. Nesse contexto, o aplicativo **Photos** utiliza técnicas de aprendizagem profunda para reconhecer rostos, objetos e cenas nas fotografias e nos vídeos armazenados no álbum do usuário e assim sugerir **slideshows** musicais personalizados (funcionalidade **Memories**). Para o aplicativo **Photos** criar automaticamente **slideshows** musicais com as imagens da Rio-2016, importamos, progressivamente, coleções de imagens com base nas classificações previamente feitas pela rede neural, usando um iPad do Laboratório VISGRAF.

Depois de preparar automaticamente o **slideshow**, o aplicativo permite personalizar algumas de suas propriedades através de seu ambiente de autoria, como adição de título e controle sobre a duração total da mídia. Além disso, em onze dos **slideshows** a trilha sonora foi sugerida pela funcionalidade **Memories**, que se utilizou de músicas oferecidas pelo serviço Apple Music.

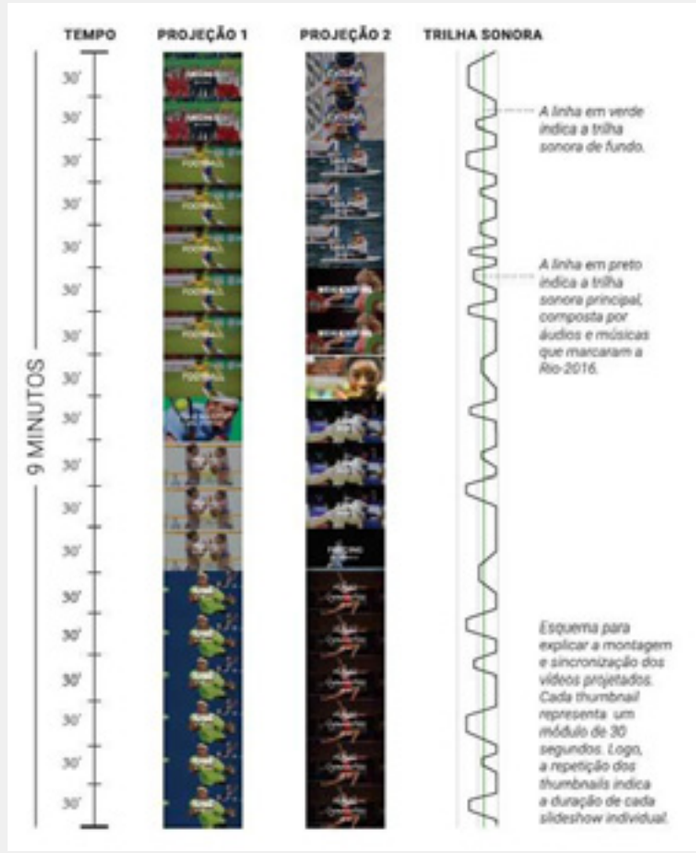


Figura **2** Montagem dos **slideshows** musicais para as projeções. Fonte: elaboração própria.

Mantivemos essas sugestões nas versões finais dos vídeos. A exceção no grupo de doze foi o **slideshow Medals**, cuja trilha sonora é uma composição algorítmica gerada a partir de uma

5 SZEGEDY, C., LIU, W., Jia, Y., SERMANET, P., REED, S., ANGUELOV, D., ... & RABINOVICH, A. (2015). Going deeper with convolutions. **In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition** (pp. 1-9).

rede complexa.[6]

Para a exibição artística no **Encontro Indisciplinas**, reunimos os doze **slideshows** musicais em dois únicos vídeos com a mesma duração de 9 minutos, para serem projetados simultaneamente. O primeiro deles agrupou, na seguinte ordem: **Tennis**, **Beach Volley**, **Juan Martín Del Potro**, **Football** e **Medals**. Já o segundo combinou, na seguinte ordem: **Olympic Gymnastics**, **Fencing**, **Judo**, **Simone Biles**, **Weightlifting**, **Cycling** e **Sailing**. Esses vídeos receberam uma trilha sonora especialmente pensada para a projeção simultânea, cujo detalhamento é apresentado na seção que se segue. O esquema na próxima página (Figura 2) ilustra a montagem dos vídeos para a projeção:

Trilha sonora

A trilha sonora para as vídeo-projeções foi criada a partir de três categorias de áudio. A primeira categoria refere-se a uma seleção de músicas compatíveis com os direitos autorais conhecidos por **Creative Commons**. Essas licenças permitem cópia e compartilhamento de músicas com menos restrições que o tradicional "Todos os direitos reservados". As músicas que foram selecionadas são instrumentais ou puramente eletrônicas. Portanto, não há vozes provenientes das músicas selecionadas na composição da trilha sonora.

A segunda categoria de áudio refere-se a uma seleção de frases marcantes que foram ditas antes e durante os Jogos Olímpicos. Muitas dessas frases foram exprimidas pelos próprios atletas em entrevistas logo após competirem e, portanto, no calor do momento. Como, por exemplo, a frase de Martine Grael após conquistar o ouro no iatismo. Enquanto outras foram manifestadas antes do início das Olimpíadas, como por exemplo a declaração da tenista Serena Williams de que era favorita. Porém ela foi eliminada logo no início do torneio de tênis. Também há frases marcantes que foram ditas por jornalistas ou por pessoas que trabalharam na organização dos Jogos, como, por exemplo, a frase de abertura dos Jogos Olímpicos, declarada pelo presidente do Comitê Olímpico Brasileiro, o senhor Carlos Nuzman.

A terceira categoria de áudio refere-se a sons marcantes das categorias esportivas que são ilustradas nas vídeo-projeções, como, por exemplo, o som da troca de bolas no tênis, o som de espadas se cruzando na esgrima e o som durante um golpe de judô, entre outros.

Como a trilha sonora foi pensada para se adequar aos dois vídeos simultaneamente, utilizou-se do artifício conhecido como **panning**, que permite que um áudio em particular, como a fala de Simone Biles, possa ser direcionado somente para uma das duas caixas acústicas (caixas de som). Nesse caso, para a caixa acústica que está relacionada com (está mais próxima) a vídeo-projeção referente à atleta Simone Biles. Portanto, durante a trilha sonora há momentos que o som é proveniente somente de uma das caixas de som. Enquanto em outros momentos o mesmo som é reproduzido nas duas caixas de som. Outra particularidade da trilha sonora é que ela foi pensada para chamar a atenção da audiência quando uma nova modalidade começa

6 ROLLA, V. G. (2015). Knowledge representation and algorithmic composition with multidigraphs. In **8th International Workshop on Machine Learning and Music (MML2015) held in conjunction with the International Symposium on Electronic Art** (ISEA2015), Vancouver, Canada.

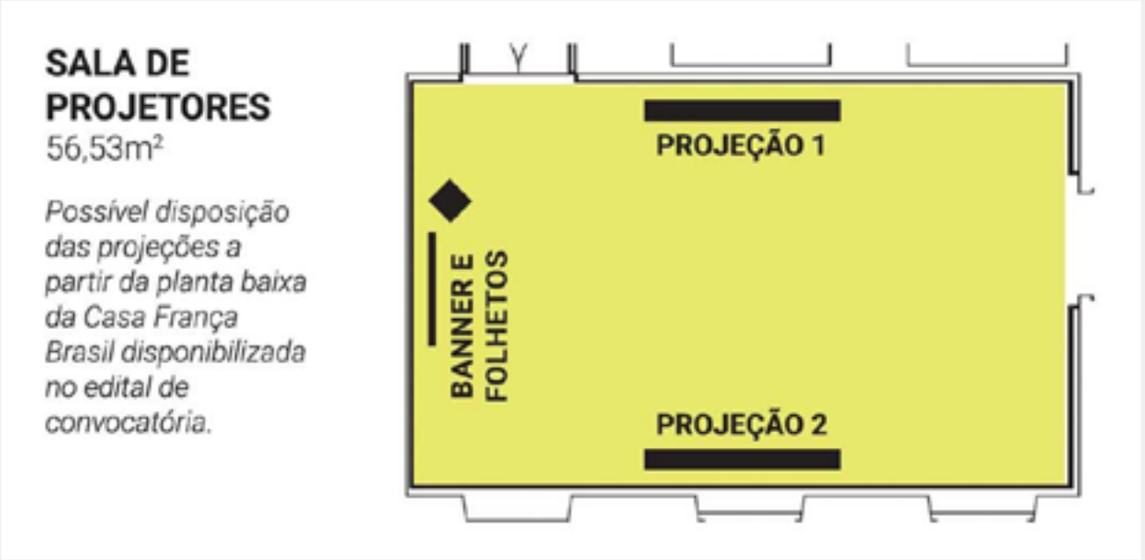


Figura **3** Projeto inicial para montagem das vídeo-projeções. Fonte: elaboração própria.

Figura **4** Parte da vídeo-projeção exibida no **Encontro** pode ser vista no Youtube. Fonte: reprodução.

a ser apresentada em uma das vídeo-projeções. Portanto, os áudios com as frases marcantes ou os sons específicos de cada modalidade esportiva, sempre marcam o começo de uma nova modalidade em uma das vídeo-projeções.

Montagem da exibição artística

No projeto inicial, os dois vídeos seriam projetados em paredes opostas (Figura 3).

Em virtude da configuração espacial da sala a nós reservada na Casa França-Brasil para a exibição artística, realizamos adaptações. As projeções foram realizadas em paredes vizinhas e deslocamos a mesa com **banner** e os folhetos explicativos para perto da porta de entrada.

As cadeiras – que ocuparam metade da sala durante as apresentações de outros trabalhos durante o evento – foram retiradas, deixando espaço para que o público pudesse circular livremente ao redor da mesa onde foram apoiados os projetores. Dispúnhamos de apenas um dos amplificadores cedidos pela organização para usarmos na sala, uma vez que os outros estavam sendo utilizados por outros palestrantes no salão principal. Como nossa trilha sonora foi especialmente preparada para ser reproduzida em duas caixas, optamos por utilizar caixas de som estéreo do Laboratório VISGRAF. Elas ficaram posicionadas próximas às paredes sobre as quais os vídeos foram projetados, bem abaixo das imagens.

O **frame** de vídeo a seguir (Figura 4) ilustra a disposição das projeções. O vídeo pode ser assistido no Youtube:[7]

Considerações finais

A convergência entre a crescente disponibilidade de grandes bases de dados visuais, avanços em técnicas de visão computacional e, em última análise, o acesso a **frameworks open-source** e bem documentados para implantação de **machine learning**, dissemina o uso de máquinas inteligentes e procedimentos automatizados na condução de trabalhos e pesquisas interessadas em problemas de análise e síntese de imagens.

Através das vídeo-projeções exibidas no **Encontro Indisciplinas** e do debate conduzido logo após a apresentação, buscamos levar aos espectadores um primeiro contato com essa abordagem de pesquisa que, dentre outros escopos, oferece uma infinidade de oportunidades para usos artísticos tão amplas quanto inexploradas.

7 O vídeo está disponível em: ←https://www.youtube.com/watch?v=kDDcKEq6U1s→.


EXPE
RI
ENC
IAS
EXPE
RI
ENC
IAS