

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Caso Eu Não Aconteça
*um estudo de realidade aumentada
aplicada ao teatro*

Gustavo Guimarães Souto

MONOGRAFIA FINAL
MAC 499 — TRABALHO DE
FORMATURA SUPERVISIONADO

Supervisor: Prof. Dr. Carlos Hitoshi Morimoto

São Paulo
2024

*O conteúdo deste trabalho é publicado sob a licença CC BY 4.0
(Creative Commons Attribution 4.0 International License)*

Quem me ama me entende

— Carlos 'Pitch' Souto

A verdade é que o amor nunca esteve sujeito ao entendimento

— Gustavo(h) 'Eu' Souto

Resumo

Gustavo Guimarães Souto. **Caso Eu Não Aconteça: *um estudo de realidade aumentada aplicada ao teatro***. Monografia (Bacharelado). Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, 2024.

Desenvolvimentos tecnológicos prestam, constantemente, possibilidades para novas composições artísticas. A utilização de técnicas de Realidade Estendida, e especificamente Realidade Aumentada, pode oferecer ferramentas baratas e escaláveis para a expansão da produção dramatológica, permitindo a inserção de efeitos especiais, objetos, elementos de cenário e auxílios sensoriais em tempo real em reação à atividade cênica, utilizando artifícios de Visão Computacional e Redes Neurais. Propõe-se criar uma dramaturgia mista, unindo elementos da programação e narratologia em prol de um processo e uma forma que adereça as necessidades de ambas.

Palavras-chave: Realidade Aumentada. Teatro. Visão Computacional.

Abstract

Gustavo Guimarães Souto. **Should I Not: a study on the application of Virtual Reality to theatre.** Capstone Project Report (Bachelor). Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2024.

Technological advancement often allow for new artistic creations. From the application of Augmented Reality techniques, it's possible to offer affordable and scalable tools to enhance dramatic performance, conjuring special effects, objects, scenery and accessibility tools that are otherwise parallel to the theatre, or can't be broadly applied for lack of investment. We propose a dramaturgy that unites elements of computer programming to theatre theory, in order to provide for the necessities of both areas.

Keywords: Augmented Reality. Theatre. Computer Vision.

Lista de figuras

2	O reflexo de um objeto fortemente iluminado é projetado sobre um vidro que separa palco e plateia. Le Monde Illustré, no. 406, 21 January 1865, p. 48	1
1	Aplicações cotidianas de sistemas de realidade aumentada	2
3	Shows holográficos ao vivo	2
4	Representação da Naumachia realizada em 57 DC, durante o reinado de Nero. Desenho por Diacomio Lauro	3
5	Experimento cênico. Um vídeo é projetado e interage com um ator (eu) seguindo um roteiro fixo. Grupo de Teatro Vaca Esférica 2023	5
2.1	Primeira configuração de palco, com tela projetiva entre público e performance.	10
2.2	Com tela montada em apenas metade do espaço, no centro do palco, voltam a existir áreas do primeiro e segundo plano de contato direto com o público.	10
2.3	Palco divide dois espaços que podem ser livremente ocupados pelo público, que ganha poder de decisão com relação a que "versão" da performance assistir	11
2.4	Frames subsequentes, gerados com variância aleatória	12
3.1	Gesto 0: Uma mão aberta	13
3.2	Gesto 1: Uma mão fechada	14
3.3	Valor F1 x Confiança para a segunda geração do sistema.	16
3.4	Precisão x Confiança para a segunda geração do sistema	16
3.5	Valor F1 x Confiança para a terceira geração do sistema.	17
3.6	Precisão x Confiança para a terceira geração do sistema	18
4.1	Valor F1 x Confiança para a ultima geração do sistema.	19
4.2	Precisão x Confiança para a ultima geração do sistema	20

Lista de tabelas

3.1	Diferentes sinais, já usando identificação por cores	15
-----	--	----

Sumário

Ethos: Introdução	1
Conceituação	1
0.1 As Origens de um Teatro Pobre e Oprimido	3
0.2 Acessibilidade	4
0.3 Cinema ao vivo	4
1 Opsis: Tecnologias / Elenco	7
1.1 YOLO - Reconhecimento	7
1.2 OpenCV - Visão	8
1.3 OpenGL - Construção Gráfica	8
1.4 Áudio Visual	8
2 Dianoia: Idealização	9
2.1 Espaço e Montagem	9
2.2 Aleatoriedade, Comunicação e o Acaso	11
2.3 Produção Mista	11
3 Mythos: Procedimento	13
3.1 Primeiro Estásimo: Construção	13
3.2 Segundo Estásimo: Cores	14
3.3 Último Estásimo: Padronização	17
4 Mellos: Resultados	19
4.0.1 Párodo	21
5 Lexis: Conclusão e Desdobramentos	23

Apêndices

A Caso Eu Não Aconteça / A Trilogia do Espaço	25
A.1 Constança	26
A.2 Dramaturgia	26

Anexos

A Código e Efeitos	29
A.1 A sequência de uma árvore	29
A.2 Filtros de cor	30
A.3 Rosto	30

Referências	33
--------------------	-----------

Ethos: Introdução

Conceituação

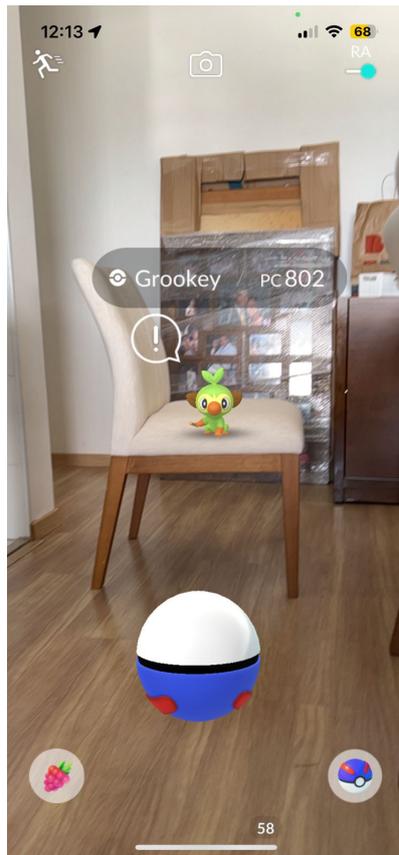
Realidade Aumentada (AR) diz respeito ao conjunto de técnicas que buscam sobrepor a experiência sensorial, complementando-a com uma dimensão artificial. Sob essa definição, pode-se empregar uma série de tecnologias diferentes, desde câmeras e *displays* vestíveis, até associações de lentes, telas e projetores. A ideia fundamental de que realidade aumentada deve ser assessorio da percepção comum, ao invés de substituí-la, é sua diferença central com relação a Realidade Virtual (VR) e outras formas de Realidade Mista (XR).

Cotidianamente, diversas formas de AR já são de acesso geral do público, seja em instrumentos de navegação como sistemas de GPS e câmeras, ou em jogos eletrônicos e equipamentos de trabalho. De fato, elementos de realidade aumentada tem sido usados pela aviação desde os anos 60, orientando muito do desenvolvimento atual com relação a essas tecnologias.

Ainda antes, porém, Realidade Aumentada já era lugar comum nas artes, figurando em efeitos especiais para filmes e apresentações ao vivo. O maior exemplo é, talvez, o chamado *pepper's ghost*, uma ilusão óptica que utiliza projeções para criar objetos e agentes no palco. O método ficou famoso pela apresentação em hologramas de artistas falecidos, como Tupac e Michael Jackson, e digitais como Vocaloids.



Figura 2: O reflexo de um objeto fortemente iluminado é projetado sobre um vidro que separa palco e plateia. *Le Monde Illustré*, no. 406, 21 January 1865, p. 48

(a) *Pokemon Go, Niantic & Pokemon Company*(b) *Câmera de ré***Figura 1:** *Aplicações cotidianas de sistemas de realidade aumentada*

Atualmente, a aplicação de técnicas do tipo, apesar de muito mais difundida, é cara e dispendiosa em um nível proibitivo para a maior parte dos artistas. *pepper's ghost* em específico exige grandes adaptações no lugar onde a apresentação é montada, o que se rende inviável para casas com apenas um palco e títulos que se alternam ao longo da semana, e mesmo sistemas digitais ainda têm que lidar com ampla quantidade de equipamentos e displays para toda a plateia, e trabalhos computacionais titânicos para efetuar todos os cálculos necessários.

(a) *Artista digital Hatsune Miku. Sega Graphics*(b) *Apresentação póstuma de Tupac, Coachella 2012***Figura 3:** *Shows holográficos ao vivo*

Nosso objetivo é, então, estudar alternativas para aplicação barata de técnicas de Realidade Aumentada às artes performáticas, com o teatro como foco.

0.1 As Origens de um Teatro Pobre e Oprimido

A performance emprega ferramentas tecnológicas avançadas há milênios. Ao ponto de que, no império romano, *Naumachias*, grandes batalhas navais chegaram a ser encenadas em anfiteatros alagados especialmente para o evento.



Figura 4: Representação da Naumachia realizada em 57 DC, durante o reinado de Nero. Desenho por Diacommo Lauro

Seguindo os avanços fabris pós-guerra, a aplicação de elementos complicados de maquiagem, cenário e efeitos se tornou cada vez mais frequente, conforme o teatro tentava emular possibilidades que se tornavam frequentes em filmes. Neste contexto, Jerzy Grotowski, diretor e ensaísta polonês, lançou as bases do que chamou de Teatro Pobre (GROTOWSKI, 1987), uma escola de produção dramática em que se pregava que elementos avulsos à atuação, como efeitos e cenários elaborados, poderiam poluir a produção cênica e prejudicar a interação do público com es atrizes e atores, que era elemento essencial. Ao mesmo tempo, e sobretudo no Brasil, o Teatro do Oprimido de Augusto Boal emprega ideias de Stanislavski e do Circo Teatro para construir alto nível de interação com a plateia que, em dados experimentos, pode entrar na cena e influenciar ativamente a direção que os acontecimentos tomam. Fica claro já em primeiro momento, então, a necessidade de que qualquer instrumento tecnológico empregado seja acessório, para que a peça possa ser aproveitada sem ele, sem prejuízo ao entendimento, para não aumentar desnecessariamente a distância entre dentro e fora do palco.

Além disso, como será explorado mais adiante, parece fundamental que a integração entre as parcelas computacional e cênica ocorra a nível de dramaturgia, concomitantemente desenvolvidas, para que as necessidades de uma não tomassem precedência sobre as da outra.

0.2 Acessibilidade

Duas questões centrais de acessibilidade norteiam, também, o projeto. Como pontuado, as tecnologias já comuns de aplicação digital e em tempo real ao teatro costumam ser proibitivamente caras para muitos dos grupos e produções que compõem o cenário performático contemporâneo. Isso exclui imediatamente a possibilidade de um sistema fortemente dependente de *displays* individuais e da necessidade de grandes alterações permanentes de palco, vez que essas soluções não poderiam ser largamente acessíveis a grupos amadores e semi-profissionais, e dificilmente seriam, também, apenas acessório destacável do texto cênico.

O segundo tópico surgiu de um desejo de design, mais que de uma especificação técnica. Na mesma proporção em que é fundamental que artistas possam comunicar suas performances sem obstruções, é necessário, também, que o público seja capaz de devidamente receber o que está sendo apresentado. Em um modelo baseado em perspectivas individuais, por exemplo, isso significa, entre outras coisas, calcular apropriadamente o ponto de vista de cada espectador, levando em consideração movimentos, variações de altura e distância para cada lugar e membro na plateia. Uma questão muito mais direta e impactante de se resolver, porém, são aspectos de sensibilidades e deficiências sensoriais.

Hoje, apenas algumas das grandes peças dispõem de tradução simultânea em LIBRAS, em dias e sessões especificamente delimitadas. Seria significativamente mais factível para grupos pequenos ter a peça inteira traduzida por intérpretes anteriormente uma única vez, para que os sinais possam ser adicionados à cena ocorrendo em tempo real a cada apresentação para textos de roteiro mais rígido, como é por exemplo o caso de musicais, ou com ao menos criação de legenda simultânea para shows de improviso, ou montagens orgânicas. A complexidade computacional é incomparavelmente menor, e seu efeito muito mais palpável em ampliar o contato entre obra e público.

0.3 Cinema ao vivo

A última preocupação relativa às especificações de projeto foi que, além de não aumentar a distância com o palco, a tecnologia fosse capaz, também, de processar todas as adições de AR em tempo real, respondendo às ações dos intérpretes. Sistemas não responsivos não são incomuns, mas limitam significativamente o que pode ser feito em termos tanto de dramaturgia quanto de improviso.



Figura 5: *Experimento cênico. Um vídeo é projetado e interage com um ator (eu) seguindo um roteiro fixo. Grupo de Teatro Vaca Esférica 2023*

Esse requisito inspira ambos, um módulo de reconhecimento que seja capaz de analisar a cada momento o estado do palco, e um sistema inteligente capaz de decidir o que fazer. É preciso que o computador esteja, de uma forma ou de outra, também em cena.

Capítulo 1

Opsis: Tecnologias / Elenco

Conhecidos os requisitos básicos da aplicação e seus princípios de execução, pode-se começar a fase de pesquisa instrumental, em que foram elencadas tecnologias para cumprir cada papel. Duas arquiteturas figuraram, principalmente, para essa escolha. Uma opção era uma rede neural treinada para reconhecer cada sinal, unida a um módulo de visão e um arcabouço de computação gráfica encadeados, mas independentes entre si; enquanto a segunda utilizaria um sistema integrado já pensado para realidade mista, sob a forma de uma aplicação em Unity.

A primeira permitiria um controle mais sutil de como estruturar cada parte dos algoritmos, e uma escolha caso a caso de cada possibilidade de subsistema, sob o risco de exigir sistemas adicionais de integração e uma maior dificuldade para a formação dos elementos visuais. A aplicação Unity, por sua vez, já vem com todo o arcabouço estruturado para uma aplicação como essa, mas exigiria o aprendizado de uma série de ferramentas novas, e uma maior complexidade de processamento.

1.1 YOLO - Reconhecimento

No que diz respeito ao módulo de reconhecimento, as ferramentas para Realidade Mista (XR) em Unity não contam com um sistema de rede neural integrado, mas com vários instrumentos para receber e processar informações obtidas pelos sensores de vários aparelhos diferentes, como câmeras, tecnologias vestíveis e celulares (TECHNOLOGIES, 2023). Indiferente da direção escolhida, este aspecto teria que ser construído à parte (KANE, 2023).

O sistema escolhido foi, então, o Ultralytics YOLO (REDMON e FARHADI, 2016, REDMON e FARHADI, 2018), especialista em tarefas de classificação, e que por utilizar uma arquitetura particional, aplica a rede neural apenas uma vez para cada imagem, processando quais regiões têm maior probabilidade de conter dado objeto. O resultado, de acordo com desenvolvedores, é a possibilidade de identificação em tempo real REDMON e FARHADI, 2018.

Todas as versões da rede neural produzida foram treinadas localmente, utilizando as arquiteturas de YOLO v8 para as versões preliminares, e YOLO v11 para a final. Os dados de treinamento advêm diretamente da produção da peça - vide capítulo 02 -, e foram anotados

utilizando Roboflow, ferramenta da própria Ultralytics.

1.2 OpenCV - Visão

Com relação ao sistema de visão, decidido a utilizar YOLO para reconhecimento, as várias ferramentas e sensores apresentadas por Unity não pareciam mais se valer de tanta utilidade. Partindo da resolução de design de que não seriam utilizadas, pelo menos para essa versão do projeto, cálculos de perspectivas individuais da plateia (como forma de atenuar os requisitos de processamento), estruturas mais simples, baseadas apenas em visualização em detrimento à riqueza que dispositivos modernos podem oferecer, e com integração facilitada à rede de reconhecimento, optou-se por utilizar do OpenCV, biblioteca de código aberto amplamente empregada e de uso já conhecido.

A função deste módulo seria captar as informações da câmera e comunicá-las à rede neural, apenas.

1.3 OpenGL - Construção Gráfica

A esse ponto, Unity já parecia cada vez menos uma opção viável, sua maior vantagem estava de fato em sua primazia gráfica. Seu sistema robusto de construção 3D permite que objetos complicados sejam construídos, e que elementos e *assets* externos fossem facilmente empregados.

Mais uma vez a escolha por simplicidade foi tomada, mas desta vez acompanhada de uma grande resolução estética. Sacrificar a precisão e riqueza de detalhes precisava incutir uma identidade visual clara que, ao invés de ser afetada pela sua natureza elementar, fazia uso dela.

Por fim, o módulo gráfico escolhido foi o OpenGL. Biblioteca com que já tinha experiência, e que pode ser utilizada em uma grande família de equipamentos, permitindo fácil adaptação caso a estratégia de representação se altere.

1.4 Áudio Visual

Os últimos sistemas a serem decididos diziam respeito à manipulação de áudio e do vídeo como entidade completa, e eram então significativamente menos vitais para o espírito do projeto. Por brevidade, os efeitos conduzidos com áudio para a formação do *Coro dos Nomes* foram processados utilizando pyDub, uma biblioteca de manipulação de arquivos de som; enquanto a manipulação de clipes para armazenamento e treinamento fez uso do MoviePy, biblioteca capaz de cortar, unir e manipular frames individuais, vídeo e áudio, além de salvá-los em diversos formatos.

Capítulo 2

Dianoia: Idealização

Detendo ambos, especificações e ferramentas, detalhes de montagem que norteariam as demais fases de produção tiveram que ser considerados. O espaço físico da montagem definiria muito do que pode e deveria ser feito em termos de efeitos de realidade aumentada, e, para que esta continuasse sendo acessório, a ambientação e atmosfera teria que ecoar com a forma.

2.1 Espaço e Montagem

Dado um espaço padrão apropriado à performance, três grandes estruturas de composição foram consideradas.

A primeira, oriunda diretamente da geometria da técnica de *pepper's ghost* tratada, coloca uma tela de projeção entre público e palco. Uma câmera grava os acontecimentos por detrás da tela, enviando as informações ao sistema de controle, que instrui o projetor onde, quando e como adicionar os efeitos de Realidade Aumentada. Esta configuração é, provavelmente a mais simples e historicamente palpável, mas tem que lidar com tanto dificuldades técnicas quanto dramatológicas.

Já em primeiro momento, a composição exige a existência de uma tela separando ator e platéia, mais do que qualquer outra, haveria o risco de aumentar a desconexão em prol dos efeitos especiais, que vai completamente contra os preceitos de teatro do oprimido e teatro pobre adotados. A solução, e de fato a forma como a técnica é originalmente aplicada, exige que a projeção se fizesse não em tela, mas em uma parede de vidro colocada à frente do palco. Isso, porém, também iria de encontro com a diretiva de que a técnica deveria ser barata para ser democratizável.

As duas composições remanescentes adereçam diretamente as falhas da primeira.

Uma solução simples é mover a tela para cobrir apenas um lado do palco, em uma estrutura mais semelhante aos experimentos assíncronos ⁵. Como resultado, pode-se apresentar em diferentes níveis do palco, incluindo à frente e atrás da tela, mas a projeção de efeitos fica restrita a metade do espaço. O movimento ganho também poderia exigir um aumento da quantidade de câmeras e com isso, sistemas adicionais (que poderiam ser

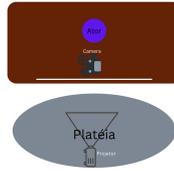


Figura 2.1: Primeira configuração de palco, com tela projetiva entre público e performance.

controlados da mesma forma que os demais, durante a peça) que informem qual imagem é captada e analisada.

Dramatologicamente, porém, esta técnica permite uma grande família de composições. A performance pode de fato lado a lado, justapondo a realidade material com a estendida; contando com movimentação constante entre regiões do palco, com as projeções manifestando agentes e entidades ativas na ação; etc.. Para a maior parte dos lugares estudados como palcos de apresentação, esta é a disposição mais facilmente aplicável utilizando infraestrutura comum.

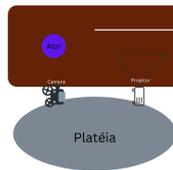


Figura 2.2: Com tela montada em apenas metade do espaço, no centro do palco, voltam a existir áreas do primeiro e segundo plano de contato direto com o público.

A última solução, e decididamente a mais inusitada, exige uma dramatização bem específica, mas é também a que mais incentiva diálogo com a platéia. Pensada para uma estrutura de teatro sem palco elevado, como a presente em ambas sedes do TUSP capital e casas de teatro alternativo da cidade, estendendo-se também a salas de aula e estúdios.

A ideia é criar um espaço especial no qual a platéia possa circular, atrás e à frente do palco. A performance é realizada em direção oposta ao principal espaço do público, organizada em uma estrutura semelhante à primeira configuração; no que seria parte reservada da parte de trás do palco, para onde o ator se apresenta, existem algumas poucas cadeiras de acesso livre, tal que o público é incentivado a escolher em cada momento assistir a versão em Realidade Aumentada, ou "entrar" para a realidade física, onde não veem nenhum efeito.

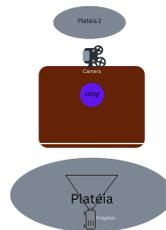


Figura 2.3: Palco divide dois espaços que podem ser livremente ocupados pelo público, que ganha poder de decisão com relação a que "versão" da performance assistir

2.2 Aleatoriedade, Comunicação e o Acaso

Sendo a leitura de cada comando realizada por rede neural, ficou claro que existe uma dimensão probabilística (ou ao menos pseudo-aleatória) com relação ao projeto. Obrigar o sistema a reconhecer e reagir aos sinais de quem está em palco trás a liberdade de improviso e o controle do tempo, impedindo que o programa possa seguir uma diretriz fixa com relação a o que, como e onde fazer.

Efetivamente, colocar o computador como reagente da ação dramática o confere papel dentro da cena, como a um ator. Claro, existe a expectativa do que faça a cada momento, mas isso se aplica igualmente a quem dá, em palco, o sinal. Ambos esperam que o roteiro seja seguido, e respondem ao tempo e chamado um do outro. Isso resvala em um aspecto essencial do Teatro do Oprimido, o contracenar como dinâmica de confiança.

Sendo o reconhecimento de deixas puramente visual, optou-se pela criação de uma série de símbolos, a ordem e o contexto dos quais responsável por comunicar, de acordo com o roteiro, o que deveria ser ativado em resposta.

Ao mesmo tempo, há o desejo de preservar diretamente parte do aspecto aleatório da arte digital algorítmica, popular ainda antes da prevalência de modelos de inteligência artificial generativa. Indo ao encontro da simplificação dos traços, fruto de utilizar sistemas gráficos menos avançados, optou-se por criar, frame a frame, qualquer objeto digital produzido utilizando funções e procedimentos que incluem fórmulas pseudo-aleatórias restritas ao escopo do roteiro (e *script*).

O resultado é um aspecto cartunesco que cria distanciamento entre a cena física e os objetos digitais, que estão sendo constantemente redesenhados de forma consistente, mas inexata.

2.3 Produção Mista

O último aspecto idealizante do sistema, é que a dimensão dramática e computacional fossem intimamente conexas, para que as necessidades de uma não conseguissem primazia sobre a outra. A comunicação simbólico-gestual para que foi decidida como indicativo de sinais e transições, exigiu da dramaturgia um formato específico, em que estes gestos

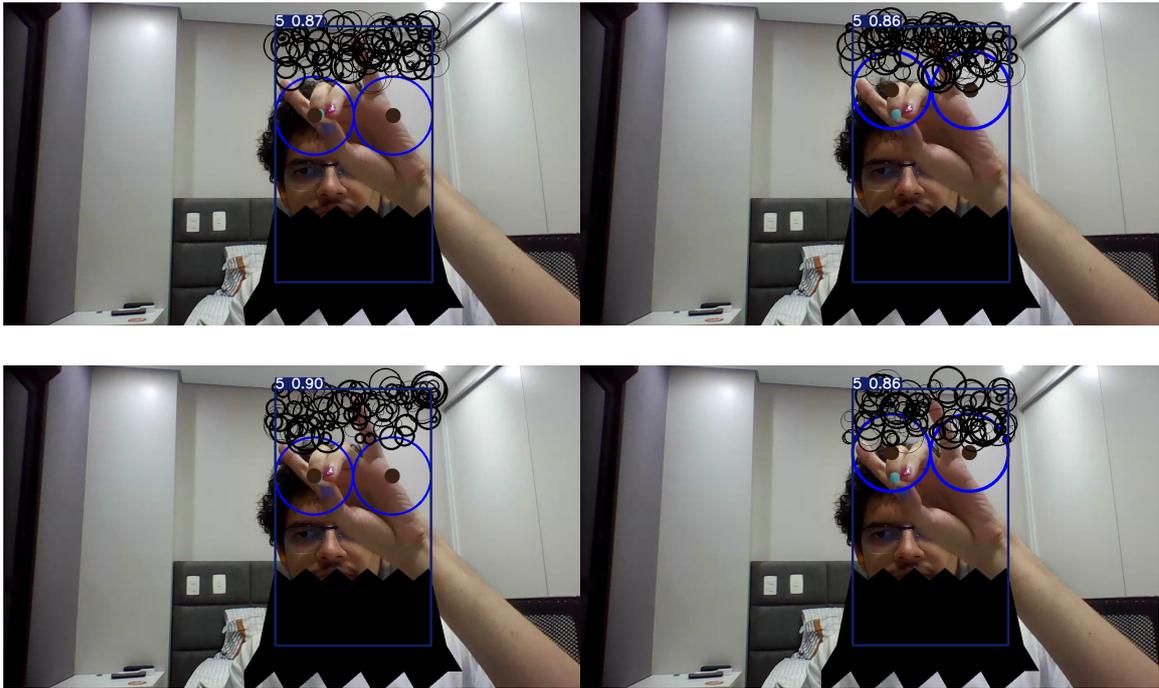


Figura 2.4: *Frames subsequentes, gerados com variância aleatória*

fossem preponderantes e fizessem sentido em meio à construção cênica. Da mesma forma, a história de aspecto mais intimista desincentiva efeitos muito chamativos e marcantes, que seriam interessantíssimos de construir.

Na versão final, apenas uma mão é utilizada, mantendo a remanescente atrás, como forma de minimizar as chances de identificar sinais equivocados. A dramaturgia, por sua vez, exigiu então que fosse utilizada a mão esquerda, e dessas duas prescrições emergiu a decisão que informa o conjunto de dados de treinamento, os sinais procurados e, é claro, o resultado tanto da eficácia narrativa, quanto da eficiência computacional do projeto como um todo.

Igualmente, a posição relativa de efeitos no enquadramento, assim como a ocupação física do espaço de apresentação andam lado a lado, sendo sucessivamente iterados ao longo de fases de produção cênica (sobretudo em ensaio) e desenvolvimento gráfico.

Capítulo 3

Mythos: Procedimento

Dada a sucessão de fases, o trabalho foi feito em etapas iterativas, tanto em termos de código quanto escrita. No geral, é possível pontuar três gerações principais da rede neural identificadora, acompanhadas pela repopulação do respectivo conjunto de treinamento.

3.1 Primeiro Estásimo: Construção

Decidido que haveriam comandos visuais determinados por gestos, um primeiro teste foi implementado para averiguar a factibilidade de reconhecer sem grades confusões o que se tentava passar. Inicialmente, o objetivo era que a rede neural pudesse discernir entre apenas dois símbolos, uma mão completamente fechada e uma mão completamente aberta.



Figura 3.1: *Gesto 0: Uma mão aberta*

Essa versão de identificador foi treinado utilizando trinta imagens, vinte e duas das quais usadas para treinamento, cinco para validação e três para teste. Cada uma pessoalmente anotada como pertencentes a uma das categorias, com a região da mão em cada foto destacada como objeto sendo procurado. Nesse primeiro momento, não houve nenhum cuidado especial com relação ao plano de fundo das imagens, ou à roupa sendo vestida.



Figura 3.2: *Gesto 1: Uma mão fechada*

Para que pudesse ser facilmente testada, essa geração foi treinada a partir da plataforma em nuvem da Ultralytics, Roboflow, que permite que uma aplicação seja exibida em browser, tanto por celular quanto por webcam.

O resultado foi promissor, apesar de modesto. O gesto de mão fechada podia ser reconhecido consistentemente a uma precisão de até 62%, enquanto a mão aberta teve confiança máxima de 47%. Estava claro, por tanto, que lidar com gestos adicionais exigiria novas formas de separação para diferenciação.

3.2 Segundo Estásimo: Cores

A solução veio por meio de uma estratégia simples e tradicional para combater a indiscernibilidade entre categorias: cores. Da mesma forma que *chroma-key* é comumente usada para aumentar o contraste de um objeto com as coisas ao redor e facilitar a identificação de limites, a hipótese era de que dedos poderiam ser mais facilmente identificáveis e diferenciáveis em cada gesto de mão se fossem marcados por cores diferentes.

Assim, para o restante dos experimentos, cada unha da mão esquerda recebeu uma cor característica e constante, tomando cuidado para que cores muito semelhantes não estivessem adjacentes, e cada configuração de sinal foi repensada para apresentar combinações diferentes de cores. Todos os sinais teriam de ser enviados utilizando o dorso da mão, em detrimento à palma como antes 3.1, e a rede neural retreinada para se ajustar aos novos símbolos.

Foram então gerados cinco sinais:

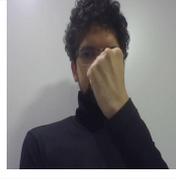
Código	Composição	Exemplo
1	Mão aberta	
2	Mão fechada	
3	Dedos anelar e médio levantados	
4	Dedos mindinho, indicador e polegar levantados	
5	Abaixados mas a mostra, dedos anelar, médio e polegar	

Tabela 3.1: *Diferentes sinais, já usando identificação por cores*

Esta segunda geração do identificador foi treinada, então, com 104 imagens referentes a todas as classes e divididas pela proporção 74 (imagens) para treinamento, 20 para validação e 10 para testes, em apenas 25 épocas e partindo da arquitetura nano-8, ainda do YOLO. A partir deste ponto, também, as redes passaram a ser treinadas localmente, vez que o sistema do Roboflow não permite o download direto dos pesos e parâmetros de rede, se computados na nuvem.

Os resultados se mostraram extremamente mais consistentes na etapa de testes do que a versão anterior, permitindo a exigência de certezas muito maiores, que seriam necessárias para a aplicação em si. Também nessa época, as primeiras versões de desenho começaram a ser esboçadas, tomando informação da posição relativa da mão em cada frame e o tamanho geral da imagem como referências, para construir principalmente elementos de cenário e filtros.

A curva que compara o valor F1 (referente a quantidade de acertos ponderada pela frequência da classe) com a confiança, revela que podemos esperar precisão mínima de 95% ao exigir certeza de 70%, revelando ganho de 23% com relação à primeira versão do sistema, mesmo com a introdução de nossas classes, apenas por aumentar o conjunto de treinamento e utilizar das cores.

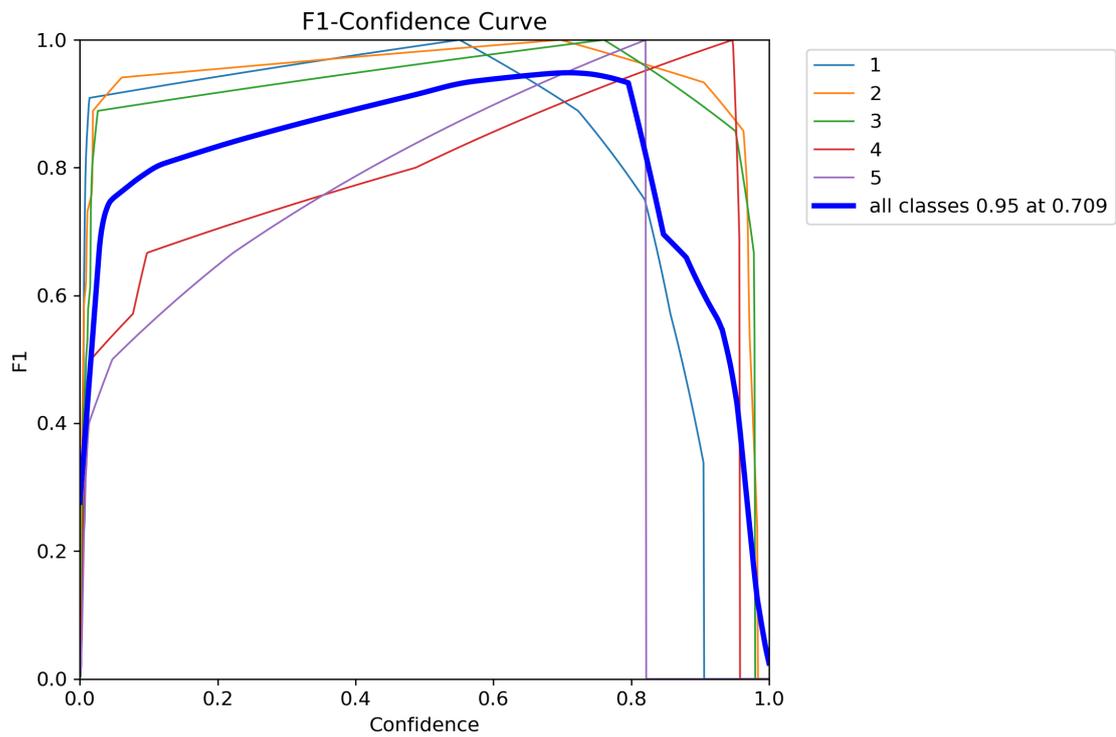


Figura 3.3: Valor F1 x Confiança para a segunda geração do sistema.

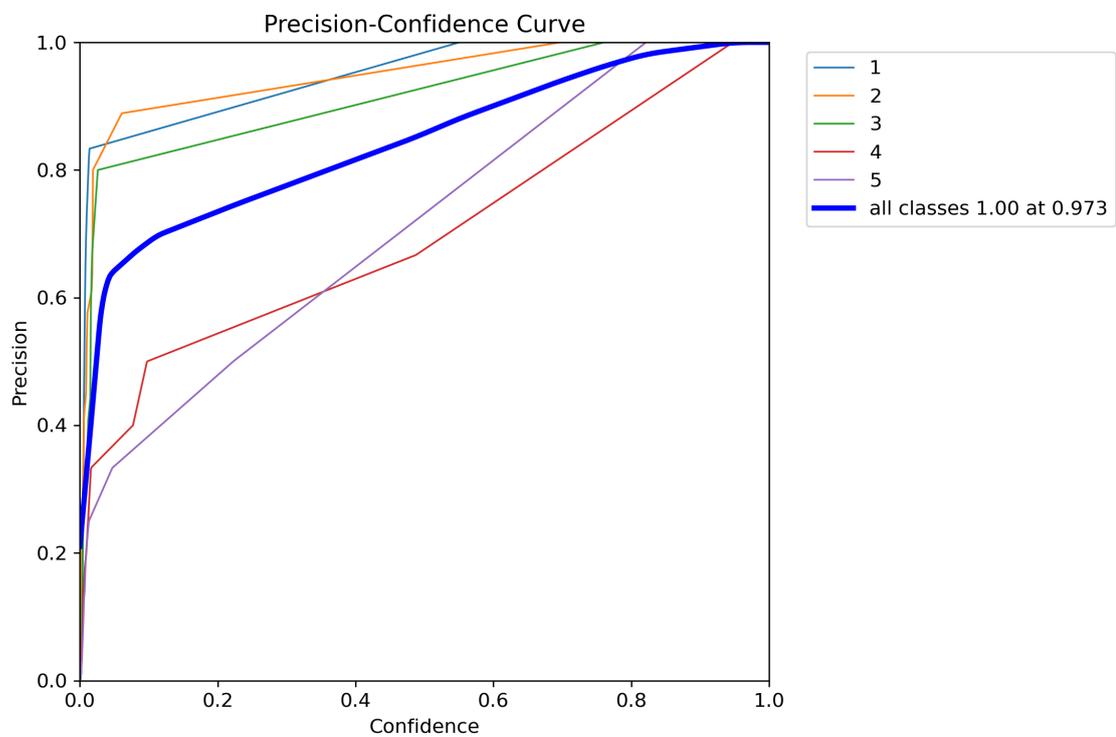


Figura 3.4: Precisão x Confiança para a segunda geração do sistema

3.3 Último Estásimo: Padronização

A última grande mudança com relação ao sistema de reconhecimento veio com duas levadas subsequentes de treinamento, a primeira utilizando fotos tiradas em ensaio, já com um fundo branco iluminado, como forma de destacar ainda mais as cores; e a segunda com gravações de ensaio. A ideia é dar peso e foco para as imagens principais a serem buscadas, representadas principalmente pelas fotos estáticas, que representam o que são chamados na animação de *key frames*.

Esta terceira geração produziu resultados mistos, se comparados à anterior.

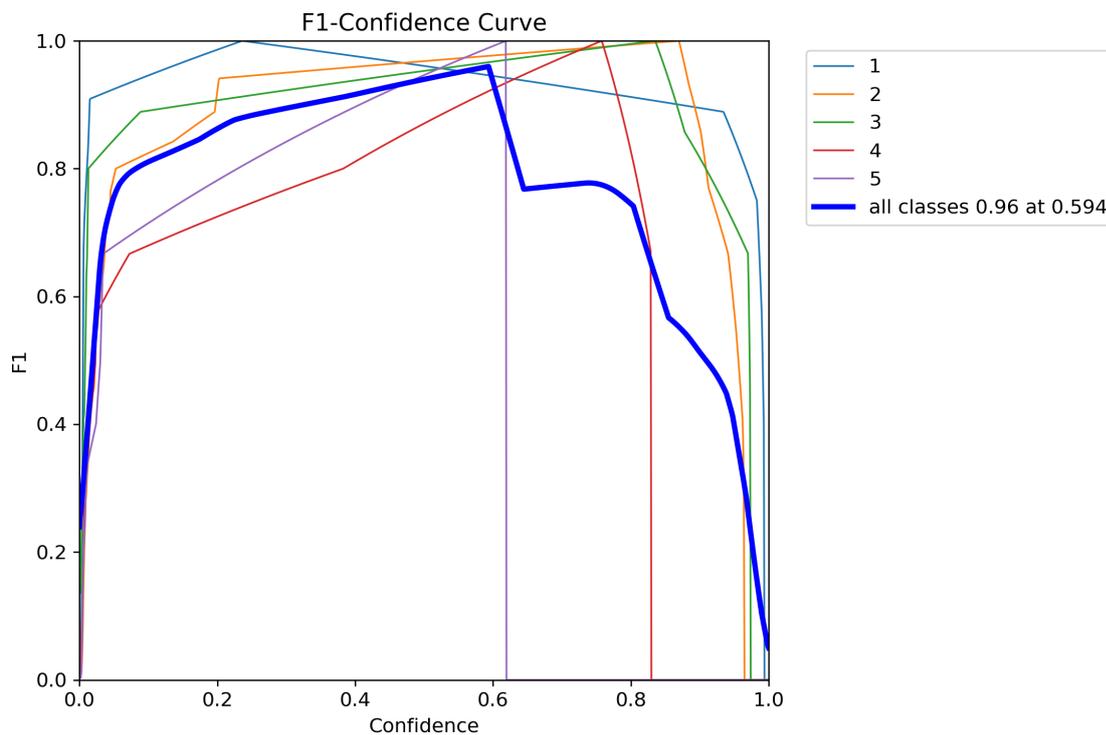


Figura 3.5: Valor F1 x Confiança para a terceira geração do sistema.

Nota-se, antes de mais nada, a piora do valor F1, que passa a alcançar 96% apenas a baixos níveis de certeza, como os registrados na primeira geração. Essa queda é, porém, observada principalmente nas classes 4 e 5, enquanto para as demais, os resultados são semelhantes aos anteriores. Ao mesmo tempo, houve ganho de precisão na identificação, quando justapostas precisão e confiança no vácuo, o que parece revelar algumas falhas com relação ao conjunto de treinamento, vez que a grandeza F1 é mais sensível à variância da proporção de cada classe na amostra. Alcançar precisão de aproximadamente 1 ao exigir confiança a partir de aproximadamente 90%, em detrimento aos 97% exigidos anteriormente, denota o amadurecimento da rede em sua habilidade de diferenciar sinais mais parecidos.

Das gravações em vídeo, por sua vez, advém imagens borradas por causa do movimento, e de momentos intermediários entre diferentes sinais, auxiliando a percepção do sistema para esses frames incertos, que causavam grande confusão no modelo anterior. Esse treinamento em vídeo também fechou por definitivo qual seria a cadeia de eventos, sinais

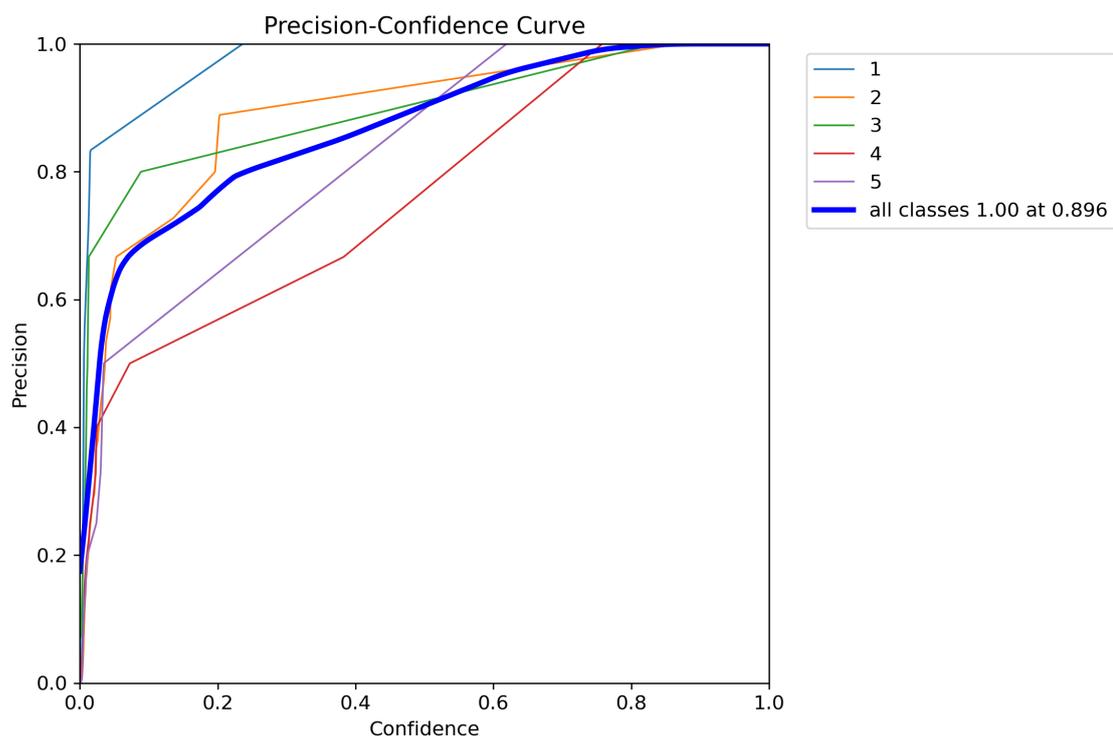


Figura 3.6: *Precisão x Confiança para a terceira geração do sistema*

e efeitos exibidos, ratificando definitivamente o roteiro.

Capítulo 4

Mellos: Resultados

A versão final do sistema de reconhecimento foi completada, então, com um corpo de 2011 imagens, sendo elas 1409 dedicadas ao treinamento, 401 à validação e 201 ao conjunto de testes, referentes a trechos de ensaios em que sinais apareciam (eliminando, assim, imagens para as quais nenhum sinal deveria ser mapeado), gravados em 24 frames por segundo. Para esta geração, foi utilizada em avanço às anteriores, YOLO v11, o modelo público mais recente, em sua versão *small*, que é a recomendada para esse nível de dados e classes. Foram performadas 291 eras, das 400 pedidas, vez que nenhuma evolução significativa fora alcançada a partir da era de número 191.

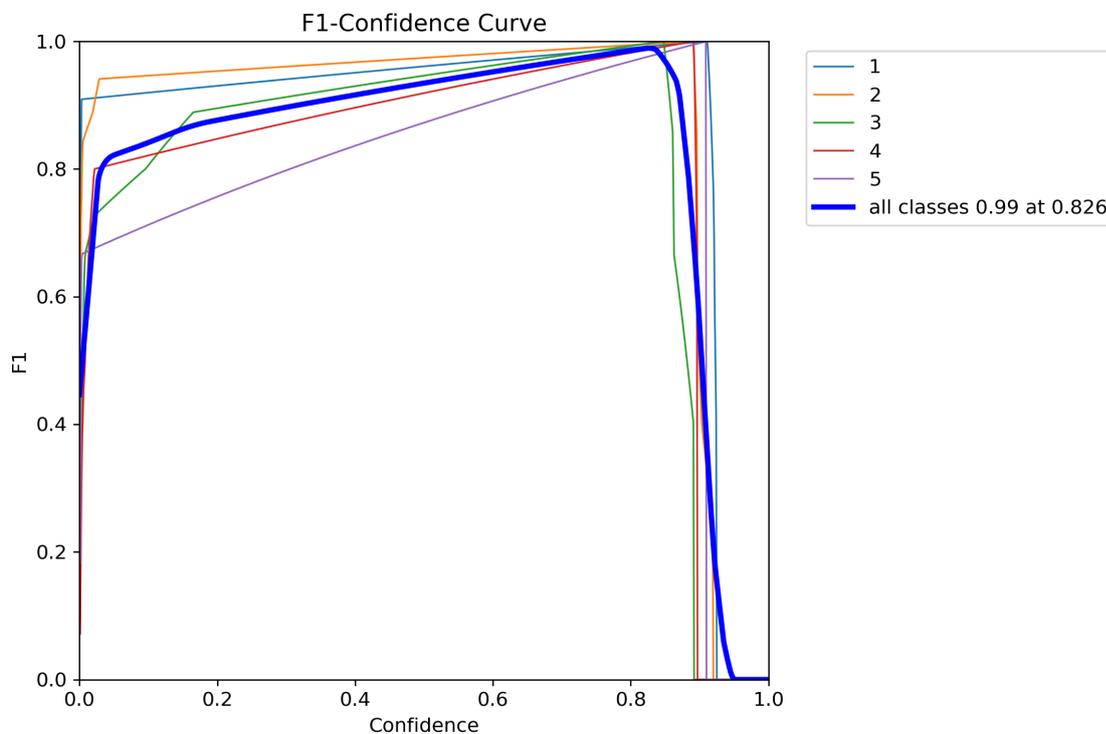


Figura 4.1: Valor F1 x Confiança para a ultima geração do sistema.

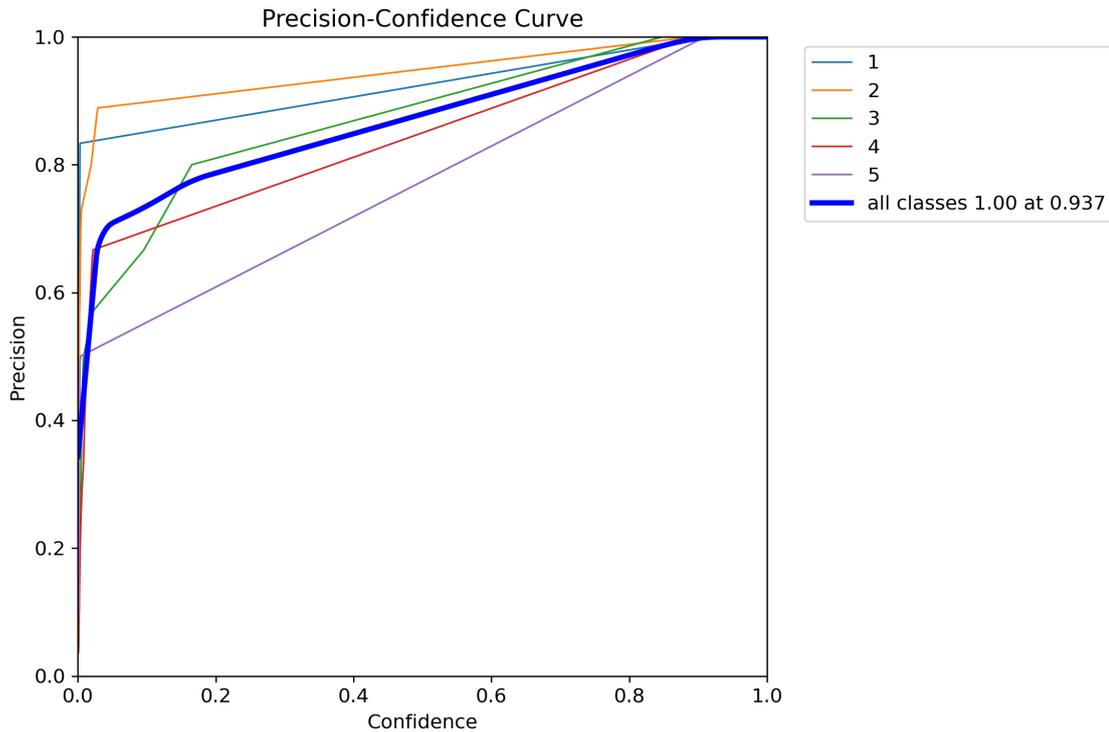


Figura 4.2: *Precisão x Confiança para a ultima geração do sistema*

Os resultados mostram melhora com relação às duas gerações remanescentes. No que diz respeito à razão entre valor F1 e confiança, pode-se esperar que 99% das asserções estejam corretas exigindo uma precisão 82%, superando o valor de 70% que era o melhor anteriormente; A comparação entre confiança e precisão caiu novamente para os níveis obtidos na segunda geração, exigindo um mínimo de 93%, pior do que a terceira versão, mas com dispersão muito menor e comportamento mais condizente entre cada classe, dando a esperar, então, que nenhum sinal seja muito menos confiável do que os demais.

Um problema que nunca pôde ser resolvido, porém, é o tempo de análise para detecção em cada frame. No computador utilizado, as primeiras gerações da rede neural levavam cerca de 0.56 segundo para cada análise, levando a uma taxa máxima próxima a dois frames por segundo, lento demais para ser processado e executado em tempo real durante a performance. Gerações posteriores melhoraram levemente esse tempo, com a versão final, baseada em YOLO 11, alcançando três frames por segundo no mesmo computador.

Diversas alternativas emergem do tratamento deste problema. A solução mais simples requer apenas o uso de uma máquina mais eficiente. O computador utilizado não possuía nenhuma placa gráfica sofisticada (particularmente efetivas para os cálculos matriciais base da rede neural), o que aumentaria exponencialmente a velocidade de processamento. De fato, uma das propostas base do sistema YOLO é a promessa de executar sistemas de visão em tempo real.

Mantendo, porém, o princípio de ser o mais democrático possível, mesmo que o investimento para um equipamento do tipo seja ordens de magnitude menores do que o necessário para equipar um teatro profissional, é possível converter o sistema para que, ao

invés de processar cada frame, possa paralelizar o esforço de aplicar a dimensão gráfica com o processamento do último sinal, tal que as imagens e efeitos projetados só sejam alterados em uma taxa levemente menor do que pode-se garantir o processamento dos sinais de vídeo.

Indiferente do método acelerar o processamento das deixas de palco, por usar sistemas de produção simples, o pós-processamento é da ordem de 30 milissegundos até para o mais complicado dos efeitos, propiciando sem grandes dificuldades a renderização ao vivo.

Para o propósito experimental deste trabalho, uma performance final ao vivo foi gravada, e depois processada pelo sistema para a introdução de filtros, efeitos de áudio e elementos gráficos, deixando a resolução desta questão em aberto, para segundo momento.

4.0.1 Párodo

Apesar de limitados em número, a associação entre os cinco símbolos obteve excelentes resultados. O sistema foi pensado para, mais do que notar os sinais em si, preservar o estado anterior e a posição geral no roteiro de efeitos, tornando a gramática que compõe dependente de contexto, e por tanto capaz de expressar uma riqueza enorme de significados, cada um podendo ser atrelada a um efeito, ou a uma combinação destes, como pode ser visto em deixas simplificadas de rubricas presentes no texto apêndice.

Principalmente em uma apresentação de grande porte, onde lê-se informação de diversos agentes diferentes, essas deixas computacionais podem se tornar muito mais sutis e complexas, integrando-se naturalmente à coreografia de cada cena.

Capítulo 5

Lexis: Conclusão e Desdobramentos

Existe uma bivalência com relação à aplicabilidade de sistemas como os estudados em um cenário teatral.

Por um lado, fica claro que as tecnologias atuais já são capazes de ser utilizadas como ferramentas adequadas para a produção artística semi-profissional ou até mesmo amadora, por um custo democrático consistindo principalmente de um computador e um par de câmeras, sem exigir modificações complexas do lugar de apresentação. Além disso, a portabilidade destas estruturas é poderosa, vez que ao desmaterializar grandes elementos de cenário, pode permitir apresentações e intervenções em ambientes menos tradicionais, levando mais peças ao grande público, e com uma equipe de produção diminuída.

Em contrapartida, os efeitos ainda foram implementados diretamente no código, dificultando a personalização destes, assim como exigindo o trabalho de alguma pessoa já acostumada com computação gráfica.

O aspecto mais impactante da produção deve ter sido, porém, a consolidação do método de produção conjunta. Muito da objetividade e algoritmidade inerentes à área da programação, incluindo pensamento lógico e aspectos normalizadores de distância e composição teve que ser levada à dramaturgia, que precisava se ocupar de comunicar claramente suas demandas. E da mesma forma, a prática computacional abriu espaço à incerteza e à estética dramatológica, envolvendo vários cenários de tentativa e erro para ajustar parâmetros dos efeitos, e até mesmo a ordem de composição de cada um. Em último nível, um se beneficiou das potências e sensibilidades do outro.

Apêndice A

Caso Eu Não Aconteça / A Trilogia do Espaço

Enquanto pensava em o que fazer como trabalho de conclusão de curso, foi um tanto difícil me livrar da ideia de que deveria ser relacionado ao teatro. De fato, muito de como interagi com a faculdade foi através de produzir e estudar artes, e parecia um tanto quanto diminutivo não adereçar isso no que deveria ser uma obra de fechamento.

Em meu primeiro ano da faculdade, ainda no Instituto de Física, tive a grande sorte de ser recebido pelo Grupo de Teatro Vaca Esférica, um coletivo de dramaturgia universitária, e foi com esse grupo que produzi muitos de meus textos e montagens. Foi por lá que tive liberdade criativa para escrever e experimentar com técnicas dramatológicas completamente diferentes do que já conhecia, com a segurança de que me apoiariam, e que eu não estaria sozinho.

A USP me deu, ainda, uma grande oportunidade de estudo teórico, por contato com a administração e curadoria do TUSP Butantã, e matérias cursadas principalmente na FFLCH, ECA e no Museu de Arte Contemporânea.

Caso Eu Não Aconteça veio, então, como sequência direta de um projeto dramático iniciado ainda em 2018. Duas peças antecedem o breve monólogo, tematicamente, e compõe o que eu tenho chamado de Trilogia do Espaço. Em primeiro momento, **Astria** conta a história de um grupo de cientistas perdido no espaço, sem saber o que fazer; é uma obra sobre proximidade, sobre aprender a confiar e se abrir para aqueles que estão (querendo ou não) ao seu redor. A sequência se deu por **Oblívio**, apresentada ainda nesse ano (2024), que conta o mistério do desaparecimento sobrenatural de uma pessoa; é uma peça sobre distância, sobre o quanto vínculos e afetos podem ser estendidos antes de romper. Parecia natural que o fim seria sobre convivência, sobre o próprio indivíduo, longe e paradoxalmente preso a si mesmo; é finalmente uma história sobre autoaceitação e sobre luto como forma de resistência.

A.1 Constança

Eu discuti muito com relação ao que deveria ser essa história. Por se tratar de um tema super pessoal, com uma temática pesada, meu primeiro instinto foi fugir para a comédia; interpolar uma série de esquetes e crônicas que juntas, contassem um panorama narrativamente maior. E então, joguei toda a ideia fora para contar a história de Constança.

Vários anos atrás, me deparei com essa garota medieval chamada Constança (VINYOLES I VIDAL, 2005). O caso judicial dela era o único relato que sobrara, e contava a história da adolescente resistindo a ser tirada de sua cidade natal e vendida como escrava. O Museu Histórico de Barcelona teve a delicadeza de me encaminhar o que tinham sobre o caso, incluindo o artigo citado, mas me debrucei muito com a questão de se poderia fazer isso, e de se sim, como contar essa história.

Constança foi uma menina afastada daqueles que amava e tudo o que conhecia, é uma história sobre distância. Mas mesmo assim encontrou forças para confiar naqueles ao seu redor e apelar para sua libertação, proximidade. É sobre honestidade consigo mesma, e sobre o desejo irresoluto de continuar sendo.

A.2 Dramaturgia

{A0: coro dos nomes} Aumentam em intensidade se sobrepondo}

{Interrupção {A1}}

(beat)

{none}

Índice:

- Introdução
- Aforismos
- Personagens
- Peste Negra
- Conclusão

Em vinte e cinco de Agosto, 1413, um caso judicial foi registrado em Sitges, Catalunha.

Hoje, Sitges é considerada a capital gay da europa. Boates, praias e bares tornam a pequena cidade um grande atrativo para turismo LGBTQIA+. Na época, nada disso existia. Com exceção dos gays. (...) E as praias, quem sabe. E Sitges era só um pequeno povoado a 30 quilômetros de Barcelona.

{A2: PRETO E BRANCO}

A grande dependência com a metrópole local fez com que os tribunais de Sitges estivessem subordinados à Catedral de Barcelona, e é nos registros desta que (a maior parte) destes autos habitam.

Eu adianto que nós não sabemos o resultado desse julgamento. (beat. {none})

É completamente factível que a garota tenha, de fato, sido vendida como escrava, e ainda narrativamente plausível que ela tenha sido liberta e vivido feliz para sempre.

Essa não é uma história sobre respostas, sobre a verdade. Sobre finais. Mas até aí, nenhuma história é.

{A0: Start tree}

Essa é, por outro lado, uma história sobre aforismos. As frases de sabedoria que ecoam por aí, e que tem a pretensão de saber da verdade. ‘É preciso uma vila para criar uma criança’ é um aforismo {A2: cresce}, conotando que crianças são responsabilidade de toda a sociedade. Mas minha categoria favorita de Aforismos são aqueles que se tratam de frases fora de contexto: “Nenhum homem é uma ilha” {A4: bloom}, “O inferno são os outros” {A1: Fall}.

Por terem só a pretensão de serem verdade, aforismos não precisam se preocupar com contradições. Ideias aparentemente opostas podem coexistir no mais perfeito caos. O que é ótimo, já que é como o mundo sempre funcionou. {A0: Fogo}

Constança, a garota que não queria ser vendida como escrava, também levava “é preciso uma vila” para outro caminho. No meio de suas súplicas, ela insistia que não fosse tirada de Barcelona, a cidade onde nasceu. Claro, a essa altura, ela não tinha mais nenhum vínculo com a cidade, mas mesmo assim, é quem a tinha criado. {A1: Blackout}

A vila de Constança era agora, um barco. E ela navegava em direção a Valência sob o comando de Antoni Rovira. {retorno}

Seus outros compatriotas incluíam duas mulheres chamadas Caterina, e Johan Boní.

Boní conhecera Constança dois dias atrás, e era ele quem a acompanhava, a força, para fora da cidade. Caterina 1 era prostituta, Caterina 2 rica.

(beat)

Os quatro eram responsáveis, diria o aforismo, por cuidar da menina, e garantir sua segurança.

Foi conversando com as Caterinas e com Antoni Rovira que Constança pode convencê-los a ir até os tribunais de Sitges para resolver a questão.

Caterina 1 estava voltando para a cidade onde morava, sua casa, e foi a primeira convencida, se não da honestidade da menina, pelo menos da importância de seu desespero. Caterina 2 visitava um familiar, e também prestou depoimento sobre o que ouvira da criança.

(beat)

Eu reforço, nós não sabemos o que aconteceu com o julgamento. Os autos não incluem uma decisão, mas mais uma vez é importante lembrar que um: esse não é o ponto; e 2: que o inferno são os outros. {A4: Rosto}

Quando Sartre constrói sua peça, ele está tentando dizer que o inferno (figurativamente) é ser julgado e comparado a outras pessoas. À parte do que isso diz sobre todo o poder judiciário, a qual tanto Constança quanto todas as demais pessoas no barco estavam sujeitas, há de se perguntar como a menina chegou nessa posição.

Sua mãe, Maria, alega Constança, era grega, o que por decisão papal, a tornava uma mulher livre. Seu pai tinha sido convocado para trabalhar para a igreja, e saiu da cidade quando ela ainda era pequena. Com a morte de Maria, a vila exigida para criar Constança estava vazia; e mais, sua salvação estava nas mãos do julgamento dos demônios que acabara de conhecer.

{End A4}

O outro ponto, significativamente menos dramático, é que isso tenha acontecido seguindo uma grande infecção da Peste Bubônica nos anos anteriores. Há de se imaginar o quanto as décadas de mortandade afetaram o imaginário popular. Mais do que nunca, qualquer um poderia estar infectado, e por projeto ou não, propagar a doença. A praga era os outros.

{A2: Filtro}

Ainda hoje, acho que um dos aspectos mais importantes de Constança é a própria presença do mito. Que 600 anos depois, alguém ainda saiba o nome dela. E que a associação com ela seja a fonte de todos os registros que existem para os outros personagens. Para essas outras pessoas. Navegadores, ricas, prostitutas... tudo o que fazem é existir em função da menina. Que seja precisa a criança.

Essa não é uma história sobre finais, mas sobre aforismos. Sobre a pretensão de saber e de ser, muito mais do que sobre a verdade.

São Paulo, 28 de novembro de 2024.

Gustavo Guimarães Souto.

{A3: Texto}

Nenhum homem é uma ilha. E eu confio absolutamente em vocês não para saber, mas pra descobrir como lidar com tudo.

Caso alguma coisa aconteça comigo, eu os amo. Muito.

E Caso Eu Não Aconteça...

Anexo A

Código e Efeitos

Seguem pequenos trechos comentados de alguns dos códigos que compõe este trabalho.

A.1 A sequência de uma árvore

Em dado momento, uma árvore é formada, floresce e tem suas folhas desaparecendo e secando, até morrer. No código, isso é feito atrás de um iterador `CreateTree`, que gera uma sequência de sua aparência ao longo do tempo. Em cada função auxiliar, `createTrunk`, `createBranches`, `createLeaves` e `createPettals`, estão os comandos sobre como produzir uma parte específica do desenho, dependendo de qual momento da vida da árvore o programa se encontra.

```

1  def createTree(canvas, advance):
2  width, height, _ = canvas.shape
3  startingTrunkColor = [20, 32, 55]
4  startingLeavesColor = 100
5  startingLeavesAmount = 200
6  treeSpace = width // 3
7  while not advance:
8      img = createTrunk(treeSpace, canvas, startingTrunkColor)
9      img = createBranches(treeSpace, img, 40, 20, startingTrunkColor)
10     img = createLeaves(treeSpace, img, startingLeavesAmount,
11                        startingLeavesColor)
12     img = createPettals(treeSpace, img, 80, [163, 36, 150])
13     yield img
14     TrunkColor = [30, 42, 65]
15     LeavesAmount = 200
16     while advance:
17         canvas = np.zeros((480, 640, 3), dtype = "uint8")
18         img = createTrunk(treeSpace, canvas, TrunkColor)
19         img = createBranches(treeSpace, img, 40, 40, TrunkColor)
20         img = createLeaves(treeSpace, img, LeavesAmount, startingLeavesColor)
21         img = createPettals(treeSpace, img, 40, [163, 36, 150])
22         TrunkColor = [TrunkColor[0] + 10, TrunkColor[1] + 10, TrunkColor[2] + 10]
23         LeavesAmount = LeavesAmount - 15
24         yield img

```

A.2 Filtros de cor

```

1  def blackWhite(canvas):
2  for i in canvas:
3      for j in i:
4          total = j[0] + [1] + [2]
5          j[0] = j[1] = j[2] = total//3
6  return canvas
7
8  moodyRed(canvas):
9  for i in canvas:
10     for j in i:
11         j[0] = int(j[0]*0.3)
12         j[1] = int(j[1]*0.3)
13
14     return canvas

```

Breves exemplos de filtros que recebem uma matriz de imagem e mudam seus valores de cor pixel a pixel para se tornar uma escala de cinza, e mais avermelhado respectivamente. A associação de várias transformações afins a essa permite uma constelação de cenários distintos.

A.3 Rosto

Um dos primeiros efeitos visuais construídos foi um pequeno rosto de marionete, com posição oscilante de olhos e cabelo, que deveria seguir a mão produzindo o sinal esperado. Dele surgiu a ideia de justapor efeitos regulares, como o da barba que sempre tem a mesma composição relativa, com os erráticos que são redefinidos a cada iteração. Um sistema como este seria o mais prejudicado por uma solução que não atualizasse o controle de sinais frame a frame, ou por uma rede neural que não fosse capaz de detectar consistentemente as mãos em posições intermediárias e borradas por movimento.

```

1  def drawFace(result):
2
3  box = result.bboxes.xyxy[0].tolist()
4  #print(box)
5  for i in range(4): box[i] = int(box[i])
6  img = result.orig_img
7  #print(f'box = {box}')
8  x = box[2] - box[0]
9  y = box[3] - box[1]
10 img = drawGlasses(box, img, x, y)
11 img = drawBeard(box, img, x, y)
12 img = drawHair(box, img, x, y)
13
14 return img
15
16 drawGlasses(box, img, x, y):
17
18 heightMod = random.randint(25, 35)
19 size = random.randint(3,5)
20 glassHeight = int(box[1] + heightMod*y/100)

```

```

21     glassRadius = int(x / 4)
22     glass1 = (int(box[0] + glassRadius), glassHeight)
23     glass2 = (int(box[2] - glassRadius), glassHeight)
24     cv2.circle(img, glass1, glassRadius, (255, 0, 0), size)
25     cv2.circle(img, glass1, x//20, (20, 32, 55), -1)
26     cv2.circle(img, glass2, glassRadius, (255, 0, 0), size)
27     cv2.circle(img, glass2, x//20, (20, 32, 55), -1)
28     return img
29
30     drawBeard(box, img, x, y):
31
32     beardHeight = int(box[1] + 0.7*y)
33     spikes = 5
34     points = calculateBeardPoints(box, beardHeight, x, y, spikes)
35     cv2.fillPoly(img, [np.array(points, dtype = np.int32).reshape(-1, 1, 2)],
36                   (0,0,0))
37
38     return img
39
40     calculateBeardPoints(box, height, x, y, spikes):
41
42     points = [[box[0], height]]
43     distortion = 0.05*y
44     dx = int(x/(2*spikes))
45     for i in range(1,2*spikes,2):
46         points.append([int(box[0] + (i)*dx), int(height+distortion)])
47         points.append([int(box[0] + (i+1)*dx), int(height)])
48     points.append([int(box[2] + 0.1*x), int(box[3])])
49     points.append([int(box[2] + 0.2*x), int(box[3] + 0.1*y)])
50
51     dbx = int(1.4*x/(3*spikes))
52     for i in range(1,3*spikes, 3):
53         points.append([int(box[2] + 0.2*x - i*dbx), int(box[3] + 0.1*y)])
54         points.append([int(box[2] + 0.2*x - (i+1)*dbx), int(box[3] + 0.15*y)])
55         points.append([int(box[2] + 0.2*x - (i+2)*dbx), int(box[3] + 0.2*y)])
56
57     points.append([int(box[0] - 0.2*x), box[3] + 0.1*y])
58     points.append([int(box[0] - 0.1*x), box[3]])
59     return points
60
61     drawHair(box, img, x, y):
62     volume = 10
63     baseColor = (0,0,0)
64     for _ in range(100):
65         center = [random.randint(box[0] - volume, box[2] + volume), random.
66                  randint(box[1] - volume, box[1] + int(0.2*y))]
67         radius = random.randint(1, 40)
68         size = random.randint(1, 4)
69         cv2.circle(img, center, radius, baseColor, size)
70     return(img)

```


Referências

- [GROTOWSKI 1987] Jerzy GROTOWSKI. “Em busca de um teatro pobre”. In: *Em busca de um teatro pobre*. Civilização Brasileira, 1987 (citado na pg. 3).
- [KANE 2023] James C. KANE. *Applied Neural Nets in Virtual Reality*. 2023. URL: medium.com/@jameskanestl/applied-neural-nets-in-virtual-reality-7381b29676a3 (acesso em 24/05/2024) (citado na pg. 7).
- [REDMON e FARHADI 2016] Joseph REDMON e Ali FARHADI. “YOLO9000: Better, Faster, Stronger”. *arXiv e-prints*, arXiv:1612.08242 (dez. de 2016), arXiv:1612.08242. DOI: [10.48550/arXiv.1612.08242](https://doi.org/10.48550/arXiv.1612.08242). arXiv: [1612.08242](https://arxiv.org/abs/1612.08242) [cs.CV] (citado na pg. 7).
- [REDMON e FARHADI 2018] Joseph REDMON e Ali FARHADI. “Yolov3: an incremental improvement”. *arXiv* (2018) (citado na pg. 7).
- [TECHNOLOGIES 2023] Unity TECHNOLOGIES. *Unity - Manual: XR*. 2023. URL: <https://docs.unity3d.com/Manual/XR.html> (acesso em 02/05/2024) (citado na pg. 7).
- [VINYOLES I VIDAL 2005] Teresa-Maria VINYOLES I VIDAL. “El clam d’una noia per no perdre els orígens, el clam per la llibertat. una estampa barcelonina del segle xv”. *Acta historica et archaeologica mediaevalia* 26 (jan. de 2005), pp. 929–943. URL: <https://raco.cat/index.php/ActaHistorica/article/view/188975> (citado na pg. 26).