

Probas de validación de Críticos de Arte Artificiais.

M^a Luz Castro Pena^a (maria.luz.castro@udc.es), Juan Jesús Romero Cardalda^a (jj@udc.es),
Penousal Machado^b (machado@dei.uc.pt), Adrián Carballal Mato^a (adrian.carballal@udc.es).

^a Facultade de Informática, Universidade da Coruña – Campus de Elviña, 15071 A Coruña,
España

^b Department of Informatics Engineering, Universidade de Coimbra – Polo II, 3030-290 Coimbra,
Portugal

Abstract — Un paso previo á construción dun artista artificial é a creación dun crítico de arte artificial. A investigación levada a cabo céntrase nos pasos iniciais dunha progresiva avaliación dos críticos artificiais. Nunha primeira fase débese identificar o autor dunha obra pictórica, e nunha segunda débese resolver un teste psicolóxico que determina as habilidades estéticas. Os resultados do CAA compáranse cos obtidos por un grupo de individuos que realizan as mesmas probas.

Index Terms — ART neural networks, Artificial Art Critics, Adaptive systems.

I. INTRODUCCIÓN

Desde o principio dos tempos a humanidade tratou de crear modelos artificiais de si mesma. Ada Lovelace, filla de Lord Byron e primeira programadora da historia, propuxo nun dos seus escritos a posibilidade de crear computadores con capacidades artísticas.

Nos últimos anos desenvolvéronse numerosos sistemas informáticos relacionados con distintas tarefas artísticas. A creación destes sistemas constitúe un grande reto, xa que a arte presenta unha serie de características (subxectiva, social, emocional) que son citadas a miúdo como propias e case exclusivas do ser humano.

Investigadores da Universidade da Coruña, da Universidade de Coimbra, e do College of Charleston veñen traballando en varias liñas de investigación relacionadas coa construción de artistas artificiais. Ademais do desenvolvemento de artistas artificiais interactivos, unha das liñas de investigación céntrase na construción de sistemas informáticos que sexan capaces de actuar como críticos de arte.

Como primeiro paso para a avaliación destes sistemas software desenvolvidos, precisamos comparar as habilidades dos críticos de arte artificiais con seres humanos con distintos coñecementos artísticos na identificación de autoría e de avaliación estética. Trátase da primeira competición home-máquina nunha tarefa que é tipicamente humana, artística e realmente complexa.

Para a realización das probas de validación ideouse un experimento en fases: na primeira, humanos e CAAs debían identificar o autor ou estilo dunha obra pictórica; na segunda debían resolver un teste psicolóxico que determina as habilidades estéticas: o teste de apreciación de debuxos de Maitland Graves [1].

Para a realización do experimento os participantes empregarán un computador sen conexión a Internet, onde se instalará un software desenvolvido para facilitar as tarefas de identificación e tratamento de datos.

A contorna de traballo do experimento é complexa, pois arte, beleza e estética son conceptos sobre cuxos lazos existen discrepancias serias na Filosofía e na Psicoloxía da Arte, desde o mesmo punto de partida da incapacidade actual para dar unha definición precisa sobre o que son. As súas relacións permiten, a pesares de todo, abordar un estudio experimental como o levado a cabo.

II. CRÍTICOS DE ARTE ARTIFICIAIS

Un Artista Artificial é, segundo a definición de Machado [2], unha aplicación capaz de crear obras con mérito estético, por si mesma ou con mínima intervención humana que debe ser capaz de aprender, mudando o seu comportamento ao longo do tempo en función do contexto e da experiencia. Ademais, para ser independente dos seres humanos, un artista artificial debe conseguir recoñecer unha obra de arte cando a “ve”.

Seguindo o exposto por Machado [3], a creación dun artista artificial require o desenvolvemento dun Crítico de Arte Artificial (CAA), que é un sistema capaz de percibir unha obra e realizar unha avaliación da mesma. O principal interese céntrase nos CAAs que producen unha avaliación cuantitativa da obra, xa que isto permite unha doada incorporación do CAA nun sistema de xeración de obras de arte.

Neste proxecto empregárase a aproximación descrita por Romero, Machado, Santos & Cardoso [4]. Esta aproximación, baseada na análise de CAAs existentes,

consiste nunha arquitectura e unha metodoloxía de validación.

A arquitectura está composta por un extractor de características e un módulo avaliador. O extractor de características é estático e específico para un dominio determinado; é o responsábel da percepción da obra de arte, xerando como saída un conxunto de métricas que reflicten as súas características relevantes. Estas métricas serven como entrada ao avaliador, que valora a obra de acordo a criterios específicos ou estéticos. O avaliador é un sistema adaptativo, o que significa que a forma en que as características son avaliadas vai mudando co tempo.

A metodoloxía de validación considera catro niveis de validación: (i) identificación de autor, (ii) identificación de estilo, (iii) avaliación estática, e (iv) avaliación dinámica. [4]. O nivel de identificación valora a habilidade do CAA para recoñecer o estilo ou a autoría dunha obra de arte. No nivel de avaliación estática a tarefa do CAA é determinar o valor estético dunha serie de pezas de arte previamente avaliadas por humanos. E no último nivel, a avaliación dinámica, os CAAs próbanse nunha sociedade híbrida composta por humanos e axentes artificiais, polo tanto o valor dunha obra depende do contexto cultural que a rodea. Con isto partimos dunha validación obxectiva (de autoría e de estilo) para despois incorporar un maior nivel de subxectividade e dinamismo, necesarios dadas as características do dominio.

III. 1ª FASE DE VALIDACIÓN: AUTORÍA

Como primeira achega á metodoloxía de validación exposta pretendemos avaliar o entrenamiento dunha RNA no proceso de recoñecemento das obras pictóricas de determinados autores, e unha das primeiras preguntas que se nos plantexan é: como recoñecemos os humanos a autoría dunha obra?, que características ten iso que denominamos “estilo” que nos permite asegurar ante unha imaxe que a obra representada corresponde a unha creación de Kandinsky?

Podemos definir o estilo como un sistema de formas, con cualidade e expresión significativas, a través da que se fan visíbeis a personalidade do artista e a perspectiva xeral dun grupo. Ademais, o estilo dun período é considerado pola maioría dos historiadores como un gusto colectivo que se fai evidente tanto nas boas como nas malas obras. [5].

Xeralmente a descrición dun estilo fai referencia a tres aspectos da arte: elementos de forma ou motivos, relacións de forma e cualidades (incluíndo unha cualidade omniabarcadora que podemos denominar “expresión”) [5]. Esta concepción de estilo xurde a partir da experiencia e a investigación. Ao correlacionar unha serie de obras de arte cun individuo ou cultura estes tres aspectos subministran un criterio máis amplo, estábel e, por tanto, máis seguro.

Mais por riba de todo isto hai que ter en conta que un artista non está presente da mesma forma en todo o que fai -aínda que algunhas características podan ser constantes-, e polo tanto un individuo pode producir durante o mesmo breve período obras que poden considerarse como dous estilos.

Para a realización deste experimento decidiuse traballar con obras de 3 autores, de forma que no experimento con humanos os participantes avaliasen 30 obras, repartidas equitativamente entre os tres autores. Precisamos un número alto de avaliacións para evitar falsos positivos debidos ao azar e tamén evitar na medida do posíbel as obras máis recoñecidas de cada un dos autores.

Escolléronse para o experimento Picasso, Kandinsky e Monet, os tres autores presentan na súa obra diferentes épocas con estilos moi distintos, sendo algunhas das súas obras moi coñecidas polo público xeral (por exemplo as obras cubistas de Picasso, os cadros abstractos de Kandinsky ou as pezas impresionistas de Monet) mentres que outras non se asocian facilmente co autor.

Escolléronse de distintas fontes de Internet un total de 666 imaxes de distintas etapas e estilos dos artistas mencionados, repartidas da seguinte forma: 212 imaxes de obras de Picasso, 339 de Monet e 115 de Kandinsky.

Para a realización do experimento con humanos desenvolveuse unha aplicación que permitise aos usuarios a realización da proba de forma sinxela e anónima, sen axuda nin intervención externa.

As imaxes dixitalizadas foron tratadas previamente, eliminando as sinaturas que revelasen a autoría das mesmas e uniformizando o seu tamaño en píxeles.

Posteriormente incluíronse nunha base de datos multimedia creada ad-hoc.

Tratouse de que esta proba fose o máis semellante ao experimento realizado coas RNAs. Escolléronse de forma aleatoria un grupo de 30 imaxes que deberán ser identificadas como correspondentes ao tipo “A”, “B” ou “C”. Aos participantes se lles entregou un libro de referencia con imaxes dos tres pintores, organizadas da seguinte forma:

Tipo A: 27 imaxes de obras de Kandinsky.

Tipo B: 27 imaxes de obras de Monet.

Tipo C: 27 imaxes de obras de Picasso.

A consulta do libro de referencia actúa como entrenamiento previo á realización do teste de identificación. Aos participantes non se lles dá información sobre o autor das obras, só se lles indica que deben encadralas dentro dun dos tres grupos, tal como fai a RNA.

Para a posterior análise dos datos das respostas é interesante coñecer o nivel de experiencia artística dos individuos, que nos permitirá clasificalos por categorías e comprobar se existe algunha diferenza na porcentaxe de acertos entre as mesmas. Ademais, solicitarase ao usuario os seus datos de idade e sexo.

Considerouse tamén de interese coñecer o resultado das respostas erradas, polo que a lista de resultados de cada individuo inclúe o número da imaxe identificada, un campo binario de acerto e o valor da resposta.

Os experimentos realizáronse na Facultade de Informática, dentro das I Xornadas UDC de Arte Artificial. Dispuxéronse 34 computadores sen conexión a Internet para a realización das probas, co programa de teste preinstalado e en execución ocupando o tamaño total da pantalla, evitando así que o usuario poda saír da mesma ou executar algún outro programa.

Antes de comezar o teste informouse aos participantes do funcionamento do programa e dos obxectivos do experimento, en ningún caso se prestou axuda durante a execución da proba.

Os participantes estiveron vixiados en todo momento para impedir o uso de material de apoio na realización das experiencias.

Ao premer o botón “Comezar” aparecen de forma sucesiva e aleatoria cada unha das 30 imaxes do grupo de teste. O usuario só ten que marcar a clasificación que considera correcta para a obra (Tipo A, B ou C) e premer o botón “Seguinte”. Baseándonos en estudos previos que teñen demostrado que o tempo de exposición non exerce influencia significativa sobre a preferencia estética [6], decidiuse non impor un límite temporal ás respostas dos participantes, e así non existe un tempo máximo marcado para a resposta a cada conxunto de imaxes, mais tampouco se permite a volta atrás. O usuario sabe en cada momento cantas imaxes leva avaliadas e cantas restan, e ao final da proba móstrase o resultado obtido.

III. 2ª FASE DE VALIDACIÓN: APRECIACIÓN ESTÉTICA

Para a realización do segundo experimento, encadrado na terceira fase da metodoloxía de validación de CAAs recorremos a un teste psicolóxico, o Design Judgement Test de Maitland Graves [1]. Este teste permite estimar certas compoñentes da aptitude para avaliar estruturas artísticas. Para isto, inténtase estimar o grao de recoñecemento e a reacción dun individuo a principios básicos de orde estética, de nome: unidade, predominancia, variedade, equilibrio, continuidade, simetría, proporción e ritmo.

O teste está constituído por 90 ítems. Cada ítem consiste nun par ou trío de deseños semellantes. Un dos deseños obedece aos principios enumerados, os restantes violan un ou máis deses principios.

A tarefa da persoa que está a ser avaliada é escoller, en cada uno dos 90 ítems, o deseño que considera ter maior valor.

Nos testes realizados polo propio Maitland Graves [1], con individuos seleccionados de forma aleatoria, resultaron nunha media de 45,680 respostas correctas; cando pasamos a alumnos de último curso de belas artes esta media sube a 55,6897. Hai que ter en conta que

responder aleatoriamente ao test da como resultado una media de 43,47 respostas correctas. A diferenza entre este resultado e o obtido por una poboación seleccionada de forma aleatoria é mínima: 2,21, o que dá unha idea da dificultade do teste.

Para o noso experimento escóllense de forma aleatoria un grupo de 30 páxinas do total de 90 que presenta o teste, que deberán ser avaliadas por todos os participantes na proba.

Antes de comezar ofrécese ao usuario unha breve descrición do teste de apreciación estética e do procedemento a seguir. Despois procédese como no primeiro experimento.

IV. EXPERIMENTOS CON RNAS

A metodoloxía utilizada no experimento con Redes de Neuronas Artificiais consta dos seguintes pasos:

A. *Obtención de datos*

A selección dos datos a utilizar para o entrenamiento das redes e a súa avaliación é unha fase moi importante. Neste caso a tarefa é complexa dado que para o primeiro teste interesa ter un grande número de imaxes de cada autor. Para conseguilo empregáronse varias fontes en Internet para crear ad hoc unha colección de obras destes tres artistas.

B. *Extractor de características*

Para a obtención das métricas correspondentes ás imaxes do teste empregouse unha versión mellorada do extractor empregado en experimentos previos [2]-[3]. O extractor de características visual está baseado na noción de que a complexidade dunha imaxe é unha característica importante para a avaliación das súas propiedades estéticas.

Como primeiro paso, o extractor normaliza as imaxes a 128x128 píxeles para evitar as diferentes relacións ancho-alto e facilitar o proceso de extracción. Posteriormente a imaxe divídese en tres canais: ton, saturación e luminosidade.

Despois de dividir a imaxe nos tres canais HSV, extráense catro familias de métricas da imaxe resultante: métricas de complexidade (fractal e JPEG), dimensión fractal, baseadas na distribución Zipf e estatísticas (media e desviación típica).

Nas métricas baseadas en complexidade estímase a “complexidade de compresión” como o ratio entre o erro xerado por un método de compresión e o ratio de compresión. (este cálculo realízase para compresión JPEG e fractal, con tres ratios de compresión en cada caso).

En relación ás métricas baseadas na Lei Zipf, empréganse dous valores por canal, correspondentes ao valor da pendente e á correlación coa pendente da distribución Zipf do valor dos píxeles da imaxe.

Para o canal de luminosidade o extractor tamén determina a dimensión fractal da imaxe orixinal e da mesma despois de aplicarlle tres tipos de filtros Sobel diferentes. Cada unha desas medidas da lugar a dous valores: a dimensión fractal e a correlación lineal coa mesma.

A dimensión fractal mediuse empregando unha aproximación semellante á que utilizou Taylor et al. [7]: a imaxe convértese a branco e negro e a dimensión fractal estímase coa técnica “box-counting”. Para calcular a dimensión fractal dos bordes empregouse un filtro Sobel para detectalos e calculouse a dimensión fractal da imaxe resultante.

Para determinar a variación das características consideradas, divídese a imaxe en cinco rexións do mesmo tamaño – os catro cuadrantes, e un rectángulo sobreposto no centro – e calcúlanse as métricas descritas para cada partición. O emprego destas subimaxes pode proporcionar información de simetría e equilibrio. Este proceso conleva un total de 246 medicións

C. Normalización de datos

Os datos de entrada que require o simulador de redes neuronais empregado, o SNNS (Stuttgart Neural Network Simulator) deben estar normalizados.

D. Definición e configuración da RNA.

A rede ten tantas neuronas na capa de entrada como características consideradas, doce neuronas na capa oculta e dúas na capa de saída (una por cada categoría). Esta arquitectura foi probada con éxito en traballos anteriores.

E. Xeración de conxuntos de entrenamento e teste.

No experimento de autoría realizáronse 30 repeticións independentes da fase de entrenamento para cada unha das arquitecturas, co obxectivo de intentar conseguir resultados estatisticamente significativos. Para cada unha desas repeticións creáronse aleatoriamente uns conxuntos de entrenamento, teste e validación que contiñan o 70%, 10% e 20% dos padróns. Empregáronse os mesmos conxuntos nas distintas arquitecturas. O entrenamento das redes de neuronas finaliza ao cumprirse un dos seguintes criterios: 1.500 ciclos de entrenamento ou un error cuadrático medio nas fases de entrenamento e teste inferior a 0,005. Estes parámetros estableceéronse empiricamente en experimentos previos.

Un dos obxectivos do experimento é a avaliación da relevancia das características integradas no extractor e a súa importancia nos resultados globais. Con esta finalidade empregáronse diferentes combinacións de características, dando lugar a diferentes RNAs con diferentes capas de entrada.

Existen combinacións que inclúen as métricas relacionadas cos tres canais de cor da imaxe, outras empregan as métricas relacionadas aos canais de

saturation e luminosidade e algunhas só consideran o canal de branco e negro. Da mesma forma, podemos distinguir as redes que inclúen características relativas ás particións das que non. Ademais, xerouse un grupo de RNAs coas diferentes características citadas mais sen ter en conta a dimensión fractal.

No experimento de valoración estética o reducido número de padróns con que contamos podería levar a resultados pouco significativos se a elección dos conxuntos de teste e entrenamento no era a axeitada. Empregouse a técnica de validación cruzada para os experimentos. Esta técnica é a máis indicada cando o conxunto de padróns é pequeno, tal como ten demostrado Goutte [8].

No noso caso, empregouse a técnica de validación cruzada con 20 conxuntos. Os padróns distribúense aleatoriamente en 20 conxuntos disxuntos dun tamaño semellante (18 deles conteñen 5 padróns e 2 conteñen 4 padróns). Así, 19 conxuntos forman parte do conxunto de entrenamento e 1 conforma o conxunto de teste. Desta forma, ningún dos padróns do conxunto de teste foi incluído no conxunto de entrenamento. Este proceso realízase 20 veces e repítese para todos os conxuntos, de forma que cada un dos padróns é incluído 19 veces no conxunto de entrenamento e unha como padrón de validación.

Os resultados conseguidos, polo tanto, comprenden todos os posibles casos e o resultado global é unha media dos obtidos para cada un dos conxuntos. A aleatoriedade na selección dos conxuntos e a realización da técnica de validación cruzada pretende evitar a manipulación ou posíbel influencia nos resultados finais, de forma que sexan o máis significativos posíbel.

V. RESULTADOS: PROBA DE AUTORÍA

No experimento de Autoría a maioría das redes encóntrase entre un 60-80% de acertos. A mellor clasificación obtén un 81,82% de acertos.

As redes que avalían a complexidade da imaxe obteñen sempre altos resultados, polo que podemos concluír que estas métricas teñen grande relevancia na consecución de resultados satisfactorios, e que forman un conxunto mínimo de métricas válidas para a aprendizaxe de RNA.

En canto aos erros de clasificación, estudando as redes que presentan os resultados máis satisfactorios, as imaxes de Kandinsky son as que acadan unha mellor clasificación, cunha porcentaxe nula de erros, e as de Picasso as que presentan unha taxa de acerto menor.

Da análise destas imaxes observamos que os erros de clasificación tenden a producirse en obras que non se corresponden co estilo “típico” do autor, chegando nalgúns casos a ser máis semellantes ao estilo doutro dos artistas avaliados.

En canto ao experimento con humanos, as probas foron realizadas a unha poboación total de 64

individuos, sendo todas elas útiles ao non presentar erros a base de datos de resultados.

Do total de individuos, 39 indicaron que non tiñan experiencia artística algunha, 23 marcaron a opción intermedia (“algunha experiencia artística”) e só 2 quedaron rexistrados como profesionais.

En canto ao sexo dos participantes, o número de mulleres foi algo inferior ao de homes (29 fronte a 35), se ben na categoría definida como *Arte1*, correspondente a “algunha experiencia artística”, a porcentaxe de mulleres (52,17%) supera levemente á de homes (47,83%).

E en relación á idade, a media sitúase nos 25 anos, aínda que na categoría denominada *Arte2*, correspondente aos profesionais, a media de idade sube ata os 45,5 anos. Isto é debido a que esta categoría só presenta 2 individuos, ambos con idades superiores á media.

A porcentaxe media de acertos é do 84,43% para toda a mostra. A diferenza entre os individuos sen experiencia ou con algunha experiencia artística márcase en 2 puntos porcentuais, sendo un 83,42% a media de respostas correctas para o grupo *Arte0* un 85,36% para o grupo *Arte1*. O grupo *Arte2* presenta unha media do 93,33%, aínda que non se pode considerar suficientemente representativo por contar só con 2 resultados.

VI. RESULTADOS: PROBA DE VALIDACIÓN ESTÉTICA

A primeira arquitectura empregada foi 120-5-2. De todas as arquitecturas probadas, esta, con 5 neuronas na capa oculta, demostrou ser a que mellores resultados obtiña. Consta de 120 neuronas na capa de entrada (as correspondentes ás 60 métricas de cada unha das dúas imaxes que compoñen o padrón). Os resultados conseguidos polo sistema proposto igualan aos obtidos por expertos en belas artes.

No seguinte experimento empréganse unicamente 20 das 60 métricas que proporciona o extractor. En concreto, utilízanse aquelas que corresponden á imaxe completa e ao rectángulo sobreposto no centro. Implementábase así unha rede con 40 neuronas na capa de entrada. O experimento pretende comprobar a suficiencia desas métricas para a execución da tarefa.

Utilizando só 20 métricas de cada imaxe o rendemento e a eficiencia do sistema aumentan de forma significativa, pasando dunha porcentaxe de acerto do 66,32% ao 70,40%.

Con este experimento demóstrase que as 40 métricas eliminadas perxudicaban a aprendizaxe do sistema, introducindo, probabelmente, variábeis que non resultan significativas para a análise da estética da imaxe.

Continuando con ese proceso de redución de métricas, o seguinte experimento emprega só 10 de cada imaxe, en concreto as correspondentes á imaxe completa,

descartando así aquelas que caracterizan de forma particular as distintas partes da mesma.

Nesta ocasión, a porcentaxe de acerto alcanza o 74,49%. Así, conseguimos un reducido número de métricas que proporcionan ao sistema os principios básicos para realizar xuízos estéticos. Esta consecuencia pode resultar de axuda na futura mellora do extractor de características.

No experimento realizado cos 64 individuos, a porcentaxe media de acertos é do 46,2% para toda a mostra. A diferenza entre os individuos sen experiencia ou con algunha experiencia artística márcase en 10 puntos porcentuais, sendo un 42,56% a media de respostas correctas para o grupo *Arte0* un 52,32% para o grupo *Arte1*. O grupo *Arte2* presenta unha media do 46,6%, aínda que non se pode considerar suficientemente representativo por contar só con 2 resultados.

VII. COMPARACIÓN DE RESULTADOS

A. Identificación

Se establecemos un gráfico comparativo entre os resultados obtidos polos humanos e os obtidos polo sistema, teríamos o que mostra a Figura 1.

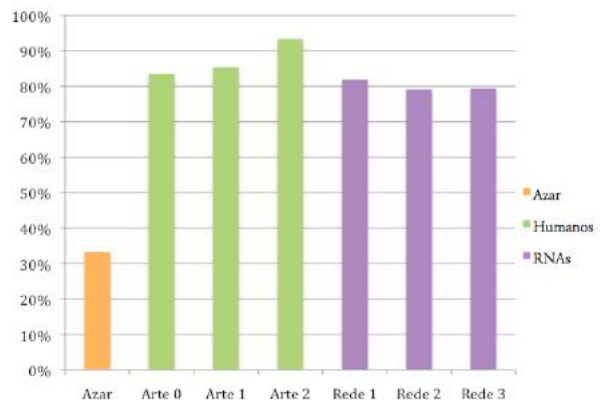


Fig. 1. Resultados humanos vs. RNAs. Proba de Autoría.

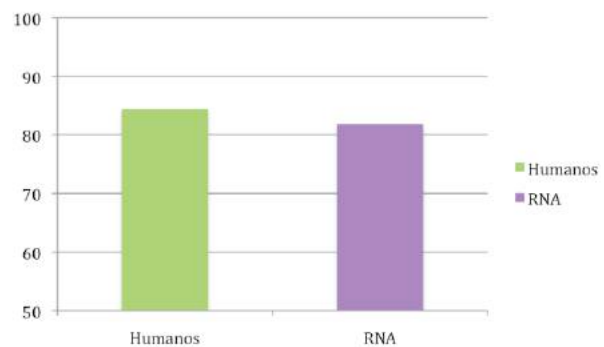


Fig. 2. Comparación de resultados. Media global. Proba de autoría.

Se comparamos a media global conseguida polos humanos co mellor resultado obtido polos sistemas, teríamos os resultados que se reflicten na Figura 2.

Estes resultados confirman a capacidade de aprendizaxe de coñecementos artísticos por parte da rede neuronal, pois o rendemento obtido polos sistemas propostos é satisfactorio e só algo inferior ao conseguido polos humanos, situándose un 5% por baixo.

Os resultados acadados polas RNAs, e en especial pola arquitectura de maior rendemento, cun 81,82% de acerto, indican que conseguiu adaptarse á tarefa proposta e, polo tanto, foron capaces de potenciar as súas habilidades de xeralización para obter as características que distinguen a autoría dunha obra. Ademais, conseguiu un conxunto de métricas que proporcionan os principios básicos necesarios para a realización desta clasificación, o que pode axudar na mellora do extractor de características visual.

B. Valoración estética.

Se establecemos un gráfico comparativo entre os resultados obtidos polos humanos e os obtidos polo sistema, teríamos o que mostra a Figura 3, sendo os sistemas A, B, e C, as arquitecturas 120-5-2, 40-5-2 e 20-5-2, respectivamente.

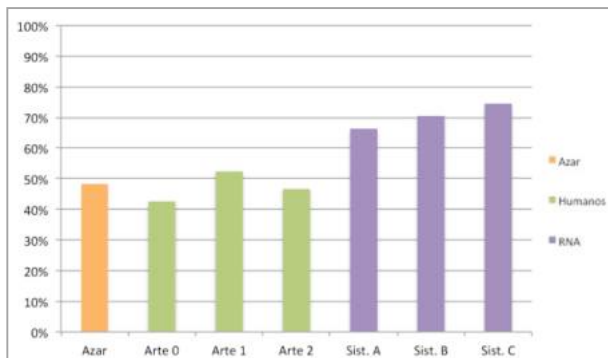


Fig. 3. Resultados humanos vs. RNAs. Proba de valoración estética.

Se comparamos a media global conseguida polos humanos coa media global obtida polos sistemas, teríamos os resultados que se reflicten na Figura 4.

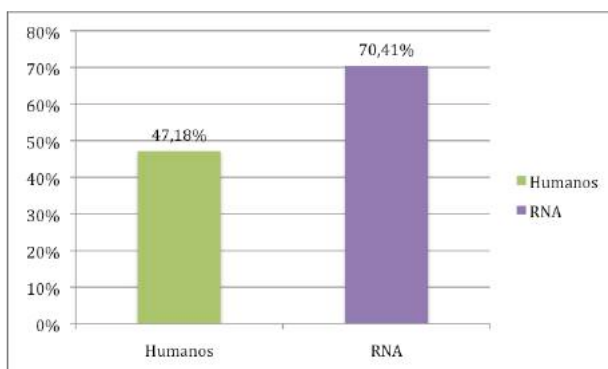


Fig. 4. Medias de resultados humanos vs. RNAs. Proba de valoración estética.

Se analizamos os valores máximos obtidos polos humanos e polos sistemas, teríamos o gráfico que aparece na Figura 5.

Analizando os resultados anteriores, o sistema sitúase un 42% por enriba, superando amplamente o sistema C (74,49% de acertos) a porcentaxe de acertos dos humanos con algunha experiencia artística (52,32%), confirmando así a capacidade de aprendizaxe de coñecementos estéticos por parte da rede neuronal. A rede foi entrenada previamente con imaxes do propio teste, mentres que os participantes no experimento non tiñan coñecemento previo do mesmo, mais un certo grado de coñecemento estético pódese presupor en todos os humanos, e especialmente se declaran ter experiencia artística.

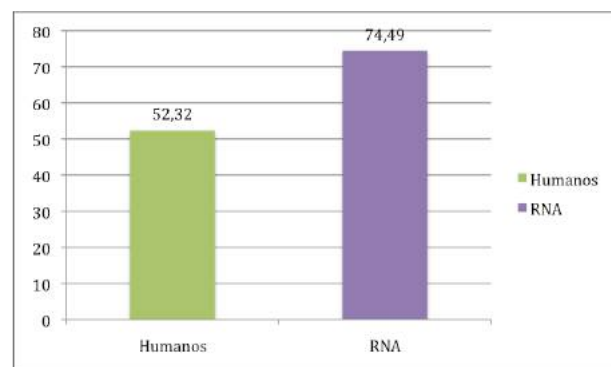


Fig. 5. Porcentaxes de acerto máximos.

Os resultados acadados por calquera dos sistemas descritos, e en especial polo sistema de maior rendemento, cun 74,49% de acerto, indican que conseguiu adaptarse á tarefa proposta e, polo tanto, foron capaces de potenciar as súas habilidades de xeralización para obter as características que distinguen unha obra que cumpre todos os criterios estéticos doutra que quebra algún deles. Ademais, conseguiu un conxunto mínimo de métricas que proporcionan os principios básicos necesarios para efectuar xuízos estéticos, o que pode axudar na mellora do extractor de características visual.

VIII. CONCLUSIÓNS

Nesta investigación experimentábase coa construción de críticos de arte artificiais para o dominio artístico visual. Para isto, séguese a arquitectura e a metodoloxía dos CAA previamente desenvolvida polo equipo de investigación e realízase un experimento que permita a valoración da metodoloxía de validación para os niveis 1 e 3.

Para a validación baseada en autoría seleccionáronse obras de tres artistas, comparando os resultados obtidos por humanos (na súa maioría alumnado universitario) cos conseguidos cun sistema de clasificación automática baseado en exemplos.

O sistema automático considera unha serie de métricas de complexidade, dimensión fractal, Zipf, así como a media e a desviación típica, en tres canais de cor (ton, saturación e luminosidade). Estas métricas alimentan un sistema conexionista entrenado a partir de exemplos. En todos os casos os resultados que se proporcionan son tomados dos conxuntos de validación e non dos conxuntos de entrenamiento.

No caso do teste con humanos mostráronse 27 obras de cada autor como referencia e solicitouse a clasificación de 30 imaxes. No caso do clasificador o conxunto empregado tiña 666 imaxes (212 de Picasso, 339 de Monet e 115 de Kandinsky). Dado que se precisan un maior número de imaxes para que o sistema poda realizar correctamente o entrenamiento, a diversidade de obras foi maior, incluíndo tanto as obras típicas como as de primeiras etapas de cada autor.

Os resultados obtidos polos humanos presentaron unha media de acertos do 84,43% (con medias do 85,36% e 83,42% para os grupos con coñecemento e sen coñecemento de arte).

Dado que se trata de 3 conxuntos, unha resposta aleatoria daría un 33,33% de acerto. O sistema automático de clasificación, empregando todas as métricas proporciona un resultado de 81,82% (3 puntos porcentuais menos que os humanos). Reducindo o número de métricas conséguense unha menor porcentaxe de acertos.

En concreto, empregando unicamente métricas de estimación de complexidade (baseada no erro de compresión fractal e jpeg), xunto coa media e a desviación típica, obtéñense resultados do 78%, moi semellantes aos obtidos con todas as métricas. Isto demostra a relevancia destas métricas para clasificacións do estilo da imaxe pictórica.

Para as probas de avaliación estética, empregouse o “Teste de Apreciación de Debuxos”, de Maitland Graves, que permite a avaliación das aptitudes artísticas das persoas. Os criterios obxectivos en que está fundamentado o mesmo permítenos unha avaliación

obxectiva nun tema tan subxectivo como é a apreciación da arte.

O sistema é sometido á realización do teste e os seus resultados son comparados cos obtidos nun experimento semellante realizado a humanos, e en todos os casos os resultados obtidos pola RNA superan aos resultados obtidos polos participantes no experimento. Non existe constancia de ningún outro traballo que probe un CAA na tarefa de avaliar esteticamente unha peza de arte.

Os resultados experimentais conseguidos demostran que as arquitecturas propostas se axustan axeitadamente ás tarefas asignadas, e que as características utilizadas son suficientes para a caracterización das obras. Ademais, a análise dos resultados permite a identificación das características máis relevantes, o que será, sen dúbida, importante para o futuro desenvolvemento dos CAAs.

REFERENCES

- [1] M. Graves, *Design Judgement Test*, The Psychological Corporation, New York, 1948.
- [2] P. Machado, *Inteligência Artificial e Arte*, PhD thesis, Universidade de Coimbra, Coimbra, Portugal, 2007.
- [3] P. Machado, J. Romero, A. Santos, A. Cardoso, B. Manaris, “Adaptative critics for evolutionary artists”, In Günther, R., et al., eds.: *Applications of Evolutionary Computing*, EvoWorkshops 2004: EvoBIO, EvoCOMNET, EvoHOT, EvoIASP, EvoMUSART, EvoSTOC. Vol. 3005 of LNCS. Coimbra, Portugal, 2004.
- [4] J. Romero, P. Machado, A. Santos, A. Cardoso, “On the Development of Critics in Evolutionary Computation Artist”, *Applications of Evolutionary Computing*, LNCS 2611, Springer-Verlag, 559-56, 2003.
- [5] M. Schapiro, *Estilo, artista y sociedad*, Teoría y filosofía del arte, Editorial Tecnos, Madrid, 1999.
- [6] M. Nadal, *Complexity and Aesthetic Preference for Diverse Visual Stimuli*, PhD thesis, Universitat de les Illes Balear, 2007.
- [7] R.P. Taylor, A.P. Jonas, *Fractal analysis of Pollock's drill paintings*, *Nature*, pp. 399, 1999.
- [8] C. Goutte, *Note on free lunches and cross-validation*, *Neural Computation*, 9, pp. 1211-1215, 1997.