

Pour une synergie entre sémiotique et *Computer Vision* : la généalogie des formes visuelles à partir de Focillon et Warburg

Maria Giulia Dondero
Fonds de la recherche scientifique – FNRS
Université de Liège
(Belgique)

L'objectif de l'œuvre d'art est de fournir la description d'une configuration conflictuelle et dynamique infiniment variée. (Thom, 2006, p. 137, je traduis)

[Dans l'art], le schéma héraclitéen de la lutte est toujours présent et même le rectangle blanc de Mondrian continue de contenir en lui la lutte entre les directions verticales et horizontales des bords du tableau. (*Ibid.*, p. 141, je traduis)

1. Introduction

Ce chapitre aborde la question de la généalogie des formes visuelles, avec comme objectif de questionner les instruments analytiques de la sémiotique visuelle et de la *Computer Vision* afin de proposer une synergie entre les deux méthodes.

Les premiers chercheurs à concevoir le problème théorique et méthodologique de la généalogie des formes visuelles, principalement artistiques, ont été des historiens de l'art, et notamment Aby Warburg avec son célèbre *Atlas Mnemosyne* (1924-1929) et Henri Focillon avec son livre séminal *Vie des formes* (1934). Les projets de ces deux historiens de l'art sont restés inachevés en raison de la difficulté à détecter des motifs similaires dans les œuvres des musées et des collections, ces dernières étant à la fois dispersées géographiquement et disparates en termes de supports et de périodes couvertes. La numérisation croissante des œuvres d'art, la disponibilité de bases de données en ligne et le traitement informatique de larges corpus d'images rendent désormais ces projets techniquement réalisables, notamment grâce à la Feature Extraction (l'extraction automatique des qualités plastiques et formelles des collections d'images) et aux techniques du Deep Learning qui s'appuient sur les réseaux neuronaux convolutifs (CNN).

J'examinerai deux projets de recherche en cours aux États-Unis et en Europe qui sont des points de départ pour réactiver les enjeux des ambitieux et exigeants projets de Warburg et de Focillon. Je focaliserai ainsi mon attention sur le projet de la *Media Visualization* du Cultural Analytics Lab dirigé par Lev Manovich, et sur le projet *Replica* du Digital Humanities Lab de l'École Polytechnique Fédérale de Lausanne afin de mettre en contraste l'approche qualitative de Warburg et Focillon et les méthodologies des analyses quantitatives et automatiques actuelles. J'essaierai également d'identifier la manière dont la sémiotique visuelle ainsi que les méthodes d'analyse automatique contemporaines peuvent apprendre de ces deux historiens de l'art qui ont été en partie délaissés à la faveur d'autres historiens de l'art plus enclins à des catégorisations définitives tel qu'Heinrich Wölfflin.

Ma proposition pour l'étude des formes visuelles sur la longue durée concerne l'analyse des *forces* contenues dans les formes. Pour repérer les répétitions et les déformations des formes, il faudra décomposer ces dernières en leurs dynamiques internes, ce que j'appelle « les forces constituant les formes ». Cette proposition devrait trouver une application dans le cadre de la sémiotique visuelle et dans le cadre des analyses automatiques. En ce qui concerne la sémiotique, il s'agirait d'associer à l'analyse classique utilisant des catégories plastiques – et, plus spécifiquement, des oppositions topologiques, chromatiques et éidétiques –, l'analyse de la *dynamique des forces*. En effet, par le passé, la sémiotique visuelle s'est consacrée davantage à analyser les oppositions plastiques que les tensions constituant les formes, en entendant par forme un effet de globalité compositionnelle en équilibre dynamique. Mon hypothèse est que l'effet de globalité est donné par des conflits tensifs au sein des compositions, que nous pouvons étudier exclusivement si nous prenons en considération les forces qui régissent les formes. L'attention doit être focalisée non seulement sur les formes visibles dans un tableau ou dans une photographie, mais aussi sur le cheminement accompli par des formes incertaines jusqu'à résolution, telles que nous pouvons en trouver trace dans les esquisses, ainsi que dans le parcours de déformation des formes d'une image à l'autre. Pour rendre compte des parcours d'une forme d'esquisse à l'autre jusqu'à l'accomplissement de l'œuvre d'art comme résultat final, et des déformations des formes d'une image à l'autre dans le cadre des reprises et des influences qui définissent la généalogie des formes, il faudrait non seulement revenir à Focillon et à Warburg pour soumettre leurs positions théoriques à l'expérimentation de la *Computer Vision* mais aussi associer cette perspective à la théorie sémiotique de la tensivité. La sémiotique tensive permet en effet de décrire les tensions et les transformations graduelles entre les termes en opposition au sein de chaque catégorie plastique (Fontanille et Zilberberg, 1998 ; Zilberberg, 2012). En vue de l'étude de la tension entre forces au sein d'une composition, la méréologie de Jean-François Bordron est aussi utile, ainsi que les propositions des mathématiques, notamment les théories des catastrophes de René Thom. Ces dernières peuvent rendre

possible la description de la composition d'une image entendue ni exclusivement comme un ensemble d'objets représentés, ni comme un jeu d'oppositions catégorielles, mais plutôt comme un ensemble de tensions en équilibre dynamique dont dépend le rythme de déploiement des formes.¹

Focaliser l'attention sur la généalogie des formes n'est pas seulement une manière de faire dialoguer l'histoire de l'art et la *Computer Vision* d'une façon renouvelée par rapport à ce qui se fait aujourd'hui dans la Digital Art History, mais aussi une manière de tenter de répondre à la difficulté qu'a la sémiotique à étudier les larges corpus et aussi les moyens corpus d'images, sur la diachronie.² Si les travaux sur l'intertextualité visuelle ont eu un certain retentissement ces vingt dernières années, l'analyse des influences d'un texte sur un autre dans le cadre de larges corpus reste problématique à résoudre avec les outils dont nous disposons.

2. Le langage computationnel et le langage visuel

Si la numérisation des œuvres d'art, la disponibilité de bases de données en ligne et le traitement informatique de larges corpus d'images rendent désormais le projet d'une généalogie des formes techniquement réalisable, pourtant, les chercheurs en *Computer Vision* n'ont pas forcément conscience des problématiques qui ont fait débat en histoire de l'art au sujet de la forme et de la déformation/transformation des formes. Ceci est bien dommage car, comme le suggère Manovich dans un article de 2020, « Computer Vision, Human Senses, and Language of Art », le langage de l'image peut bien être assumé par la logique computationnelle. Cette dernière permet d'étudier les moindres différences entre des images très similaires ainsi que des convergences et des similitudes entre images qui, à première vue, semblent n'avoir rien en commun. L'ordinateur peut voir des différences et des similitudes là où nous ne les voyons pas. En outre, comme l'affirme Manovich, les transformations rythmiques sont facilement discernées par la machine. En particulier, le langage de la description mathématique peut être utile à la sémiotique visuelle et à l'histoire de l'art, car la description par des nombres est plus fidèle que celle obtenue par les moyens de la description verbale.³ Dans l'article cité, Manovich affirme qu'aucun langage ne peut

¹ Voir à ce propos la proposition de Massaert (2019).

² Une première tentative avait été faite par Floch (1985) dans son analyse de l'intertexte de la *Composition IV* de Kandinsky. Pour une analyse complète et exemplaire de l'ensemble de la production du photographe Denis Roche, voir le chapitre « Photos en forme de "nous" : l'éclipse représentationnelle d'un couple » dans Basso et Dondero (2011).

³ « Les caractéristiques numériques de l'image qu'on utilise dans le *Computer Vision* nous offrent un langage capable de représenter des transformations temporelles graduelles et continues » (Manovich, 2020, en ligne). Sur l'analyse des rythmes des formes dans les images fixes, je renvoie le lecteur à Fontanille (1995, 1998), Colas-Blaise (2019) et Dondero (2020a).

représenter un si grand nombre de couleurs telles que celles contenues dans les formats numériques des images⁴ et qu'

[...] une représentation numérique d'un artefact culturel ou d'une expérience qui utilise des valeurs sous forme de nombres ou des caractéristiques qui ont été implémentées à partir de ces valeurs peuvent capturer des dimensions équivalentes aux dimensions de ces artefacts, avec une plus grande précision que les descriptions linguistiques. (2020, en ligne, je traduis)

La sémiotique greimassienne s'est d'ailleurs longuement opposée à la sémiologie barthésienne et benvenistienne justement à ce propos : la sémiologie voit dans le langage verbal un traducteur global et un interprète universel de tous les autres langages (visuel, audiovisuel, musical, gestuel), tandis que la sémiotique reconnaît à tout langage la capacité de s'auto-décrire et de décrire les autres langages. Les critiques faites à la sémiologie barthésienne par Paolo Fabbri dans *Le tournant sémiotique* (1998) sont bien connues.⁵

En sémiotique visuelle, l'image a été décrite non seulement dans sa dimension figurative, mais aussi dans sa dimension plastique, comme affirmé précédemment. La dimension figurative est encore redevable de la reconnaissance d'objets et du classement du monde selon des lexicalisations qui n'ont rien en commun avec l'organisation compositionnelle du langage visuel. L'analyse plastique, en revanche, dans les années 1980, a signifié une révolution dans le cadre des études du visuel en proposant une méthodologie d'analyse de la dimension plastique et en particulier des catégories eidétiques, chromatiques et topologiques – qui n'ont aucune expression lexicale équivalente dans le langage verbal.

À présent, dans la *Computer Vision*, la méthode de la « *Feature Extraction* », à savoir l'extraction des qualités plastiques des images, utilisée par Lev Manovich dans ses visualisations d'images, ainsi que des méthodes récemment réadaptées, telles que le Deep Learning utilisant les réseaux neuronaux convolutifs (CNN) (Le Cun *et al.*, 2015) – utilisé dans le projet *Replica* (Seguin, 2018) –, permettent de reconnaître la répétition des mêmes motifs dans de très grands ensembles de données d'images sans passer par une analyse des images par métadonnées. Les métadonnées ne sont en effet pas utiles pour comprendre les similarités entre les images – comme déjà remarqué à propos de la différence entre sémiologie et sémiotique. La date, l'auteur, la thématique, la liste des objets représentés, le motif iconographique ne permettent pas d'identifier la particularité et la singularité d'une image car les métadonnées rendent l'image dépendante d'un classement tout à fait étranger à ses propres fonctionnements. En effet, les méthodes computa-

⁴ « Aucun langage ne peut représenter autant de couleurs que les formats numériques de l'image » (Manovich, 2020, en ligne).

⁵ Je ne veux pas m'étendre sur la distinction *sémiotique/sémiologie* car je l'ai déjà abondamment traitée (voir Dondero, 2017 et 2020a).

tionnelles doivent pouvoir nous montrer quelque chose qui n'est pas immédiatement visible et qui n'a pas encore été classé ou reconnu par les historiens de l'art qui utilisent des méthodologies d'analyse classiques. C'est la raison pour laquelle les méthodes computationnelles sont les plus intéressantes pour l'avenir de l'analyse plastique des larges corpus d'images et donc pour l'analyse de l'histoire des formes dans toute leur irrégularité et imprévisibilité.

Je ne souhaite pas entrer dans les détails de ces deux méthodologies, mais tout simplement rappeler que les résultats de la Feature Extraction sont encore en partie redevables des choix des catégories des chercheurs, tandis que les résultats obtenus par les analyses par Deep Learning dépendent davantage de l'apprentissage que la machine aura accompli à partir des expériences obtenues sur des « training cases ». En ce sens, le Deep Learning apparaît comme laissant plus d'espace au calcul de la machine sur les images et moins aux catégories préétablies par les chercheurs. Une autre différence entre les deux méthodes concerne les volumes de données qu'elles peuvent prendre en compte : ce que l'on appelle les Big Data dans le cas du Deep Learning, et des volumes de données plus limités dans le cadre de la Feature Extraction.

Ce que ces deux méthodologies ont en commun et que les deux projets, la *Media Visualization* et *Replica*, partagent, est le choix de la visualisation comme instrument utile dans l'analyse des corpus d'images.

La *Media Visualization* permet d'analyser la production d'un designer ou d'un peintre à travers un diagramme d'images, voire un diagramme, qui organise les images selon leurs similitudes plastiques et formelles :

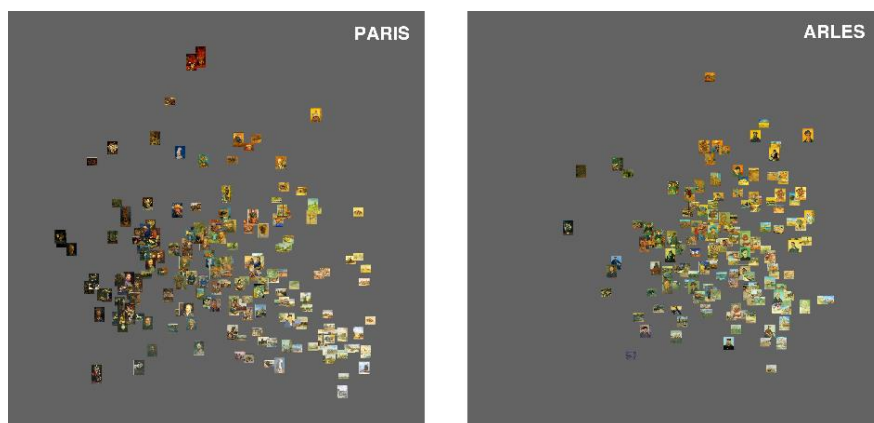


Figure 1. Comparaison entre les tableaux produits par Van Gogh à Paris (gauche) et Arles (droite) selon les paramètres de la luminosité (axe X) et de la saturation (axe Y)

Ces diagrammes d'images (Manovich, 2011) utilisent des catégories plastiques pour effectuer une classification et un groupement des éléments du corpus via la Feature Extraction, ce qui constitue en soi un processus d'analyse

(division).⁶ Cette analyse doit être comprise comme le résultat d'opérations métréologiques de séparation, de regroupement, de superposition, effectuées par la machine. Dans la figure 1, les tableaux de Van Gogh qui partagent les mêmes caractéristiques sont situés dans les mêmes zones du diagramme, tandis que celles qui ne lui ressemblent pas sont placées dans des zones opposées, à savoir des zones caractérisées par des images ayant des qualités visuelles différentes des images regroupées dans les autres zones. La position de chaque image au sein d'un groupe d'images offre une caractérisation précise de ses propriétés plastiques et les met en contraste avec les qualités d'autres groupes d'images. La représentation spatiale de l'intensité de chaque paramètre de classification (saturation, luminosité, types de contours, angles, etc.) est régie par des axes d'abscisses et d'ordonnées.

Dans le projet *Replica*, porté par Benoit Seguin (2018), la visualisation analytique des images, appelée « morphographe », montre également les relations entre des groupes d'images – dans ce cas, appartenant aux collections de la Fondazione Cini de Venise :



Figure 2. *Replica* Project, Digital Humanities Lab, EPFL (Seguin, 2018)

Les connexions visuelles entre images qui se ressemblent sont appelées *relations morphologiques* et mettent en valeur la *propagation de motifs iconographiques*.

Dans ces visualisations automatiques, on peut facilement reconnaître l'héritage des planches d'images de Aby Warburg, publiés dans l'*Atlas Mnémosyne* (2012) :

⁶ Pour un approfondissement sur ce genre d'analyse, voir Dondero (2017, 2019).



Figure 3. *Bilderatlas*, Panel 45 (Warburg, 2012)

En tant que premiers efforts de visualisation des images sur la base de leurs similitudes en termes de formes et de dynamique gestuelle et passionnelle des figures représentées, ces planches représentent une tentative très importante pour comprendre l'influence des images sur les images, par le biais d'un support visuel, le panneau fournissant une surface aidant à appréhender la relation graduée entre les images visuellement proches ou éloignées.

Warburg a été le premier à proposer une exemplification visuelle, via des planches d'images, des similarités formelles en exploitant le potentiel de la visualisation d'images. Ce genre de visualisation rend *perceptivement compréhensible* la migration de motifs d'une image à une autre. Pour cette raison, Warburg peut actuellement être considéré comme le précurseur de la *métrique de similarité* entre une image et une autre, métrique qui est mesurée à présent en Digital Art History par des algorithmes.

En revenant à la Media Visualization et à Replica, les critiques que l'on peut faire à ces deux projets sont notamment au nombre de deux : 1. la première concerne la *segmentation de l'image*. Le projet des Cultural Analytics ne prend en considération que les caractéristiques plastiques *quantifiables* des images (par exemple l'intensité lumineuse), mais pas la composition globale, ce que j'appelle la « méréologie » de l'image artistique (la relation entre la totalité et les parties). Chaque image est en effet caractérisée par une *moyenne* (chromatique, lumineuse, etc.) qui ne rend pas compte de sa complexité et de ses dynamiques internes. Le projet *Replica*,

quant à lui, vise surtout à repérer les motifs iconographiques qui se répètent d'image en image (la Vénus allongée, la femme accroupie, etc.), et porte peu d'attention à la composition globale et aux rythmes des images. 2. La seconde critique est que le travail sur les similitudes visuelles a comme objectif la description de la reprise de formes sur des *périodes assez courtes*, ce qui ne permet pas de véritablement envisager une étude sur la longue durée – contrairement à ce qui était envisagé par Warburg.

Face à ces deux projets, il est utile de revenir aux connaissances fondamentales développées en histoire de l'art sur la généalogie des formes, pour tenter de développer une analyse automatique renouvelée des grandes collections d'images.⁷

3. La forme selon les conceptions de Focillon et Warburg

Les travaux de Focillon et de Warburg proposent tous deux une conception *biologique* des formes, en plus de partager une vision similaire de la *succession des formes dans le temps*.

La première notion importante que Focillon et Warburg ont examinée est l'intervalle entre la manifestation d'une première forme et la manifestation d'une autre qui prend cette première forme et la transforme. Concernant l'intervalle, Focillon affirme :

Ce n'est pas tout de savoir que les faits se succèdent, ils se succèdent à de certains intervalles. Et ces intervalles mêmes n'autorisent pas seulement une mise en place, mais déjà, sous certaines réserves, une interprétation. *Le rapport de deux faits dans le temps n'est pas le même selon qu'ils sont plus ou moins éloignés l'un de l'autre*. Il y a là quelque chose d'analogue aux rapports des objets dans l'espace et sous la lumière, à leur dimension relative, à la projection de leurs ombres. (1934, p. 82, je souligne)

Ce que, dans *Vie des formes*, Focillon appelle génériquement « métamorphose », Warburg l'appelle plus techniquement un *processus de « dépolarisation » et de « repolarisation » des forces au sein des formes visuelles*. Pour Warburg, qui vise à construire une « iconologie de l'intervalle », l'intervalle est chargé d'énergie prête à être libérée et repolarisée. De plus, les deux chercheurs ont travaillé sur les effets latents des formes. À ce sujet, l'une des phrases les plus suggestives décrit l'émergence d'une forme à partir d'une autre, et ce en ayant recours à la notion de halo, au sens d'« aura » :

C'est que la forme est enveloppée d'un *halo*. Elle est *stricte définition de l'espace mais elle est suggestion d'autres formes*. Elle se continue, elle se

⁷ D'autres critiques sont adressées à la recherche actuelle en Digital Art History et aux Digital Humanities. Voir par exemple Hall (2013) et Bishop (2018). Ces critiques concernent la primauté de la méthodologie sur la théorie, ainsi que la primauté du positivisme mécanique sur la pensée critique.

propage dans l'imaginaire, ou plutôt nous la considérons comme une sorte de fissure, par laquelle nous pouvons faire entrer dans un règne incertain, qui n'est ni l'étendue ni le pensée, *une foule d'images qui aspirent à naître.* (*Ibid.*, p. 6, je souligne)

Dans ce sens, l'image est valorisée par Focillon comme, d'une part, quelque chose de très stable et déterminé, identifiable par « une stricte définition de l'espace » et une *évidence bien définie*, et, d'autre part, comme quelque chose qui construit un réseaux d'expériences, certaines passées, d'autres à venir, qui font partie de son aura, de son halo, à savoir de toutes les images qui l'ont influencée et qu'elle influencera.

Une autre similitude entre la pensée de Warburg et celle de Focillon concerne la conception du conflit et la notion de stratification des formes de différentes époques au sein d'une même période de l'histoire. Aby Warburg identifie les conflits actifs dans le style de la Renaissance, marqué par la coexistence de la pensée apollinienne et de la pensée dionysiaque (pensons à la figure de la nymphe, par exemple). À ce sujet, nous pouvons rappeler que Warburg focalise longuement son attention sur le tableau de Ghirlandaio, la *Naissance de Jean-Baptiste* (1485-1490, Florence). Ce qui attire l'attention de Warburg est la présence d'une « mémoire figurative » qu'il reconnaît comme appartenant au paganisme (la figure de la servante portant la corbeille de fruits) et donc l'intrusion d'une représentation de la religion payenne au sein d'un contexte chrétien. La silhouette révélerait son origine : l'équilibre, la robe et le geste avec lequel elle porte le panier de fruits sur sa tête mettent en évidence le mouvement qui caractérise la ménade. La jeune femme entre dans la scène de la nativité comme si elle venait d'un autre endroit qui est en fait une autre époque de l'histoire. L'intérêt de cette figure consiste ainsi dans le fait qu'elle conduit au problème du temps dans l'histoire de l'art et, plus précisément, à la dimension de l'anachronisme.⁸

Focillon est très proche de cette idée warburghienne, notamment lorsqu'il affirme que l'histoire doit être comprise comme une « superposition de présents largement étendus » (*Ibid.*, p. 84). Cela signifie que pour Focillon aussi, des caractéristiques formelles et sémantiques hétérogènes peuvent coexister au sein d'un même style à un même moment de l'histoire :

Du fait que les divers modes de l'action sont contemporains, c'est-à-dire saisis au même instant, il ne s'ensuit pas qu'ils soient tous au même point de leur développement. [...]. Chaque ordre de l'action obéit à son mouvement

⁸ Didi-Huberman revient sur la coexistence de plusieurs formes telle qu'elle a été pensée par Warburg. Autrement dit : certaines formes sont chargées d'une force qui vient d'ailleurs : « C'est l'étrangeté qui, dans la collision anachronique du Maintenant (la servante) et de l'Autrefois (la Victoire), ouvre au style son futur même, sa capacité à changer et à se reformer entièrement – comme Warburg l'énonce quelquefois sous le terme d'*Umstilisierung* » (2001, p. 2).

propre, *déterminé par des exigences intérieures, ralenti ou accéléré par des contacts*. Non seulement ces mouvements sont dissemblables entre eux, mais chacun d'eux n'est pas uniforme. L'histoire de l'art nous montre, juxtaposées dans le même moment, des survivances et des anticipations, des formes lentes, retardataires, contemporaines de formes hardies et rapides. (*Ibid.*, je souligne)

Prenons cet autre passage de Focillon pour illustrer ce propos qui peut paraître abstrait : « Le gothique anglais est longtemps fidèle à la conception des masses de l'art normand, alors que, dans le tracé des courbes, il anticipe avec rapidité, étant ainsi à la fois et dans le même temps un art précoce et un art conservateur » (*Ibid.*, p. 90). Focillon et Warburg partagent ainsi l'idée qu'il est possible que les parties d'une image ne soient pas synchrones.

3.1. Formes et forces

La tension entre l'évidence de l'œuvre d'art, son unicité et le fait qu'elle réalise une totalité de sens accomplie, et la capacité de cette dernière à générer d'autres images via l'aura qu'elle crée tout autour d'elle – qui fait que chaque chef-d'œuvre est continuellement réinterprété par d'autres images –, soulève la question de l'esquisse qui l'a précédée. L'esquisse attire l'attention sur le fait que chaque œuvre d'art doit être étudiée à travers les mouvements formels qui l'ont préparée. En effet, la version finale d'une image ne constitue qu'une des perspectives d'investigation possibles, tandis que la perspective des esquisses permet de prendre en compte les *oscillations* qui ont préparé la forme définitive. Focillon affirme en effet que l'esquisse « fait bouger le chef d'œuvre. Vingt expériences, *récentes ou prochaines*, entrelacent leur réseau derrière l'*évidence bien définie* de l'image » (*Ibid.*, p. 10, je souligne).

Cette multiplication des images relatives à une même œuvre permet d'analyser les formes de chaque image du point de vue de son histoire de production, ce qui a été négligé par la Digital Art History.⁹

La perspective de l'esquisse conçoit la forme comme une tension variable entre des forces contrastées/opposées, telles qu'elles ont été décrites par Deleuze dans son livre séminal *Francis Bacon : la logique de la sensation* (1981). Les peintures de Bacon sont construites sur la base de forces de pression et de dilatation, contraction, aplatissement, étirement sur des corps qui dépassent leur cadre et qui, par des sortes d'acrobaties, sont compressées et absorbées par des dispositifs tels que des éviers, des toilettes ou des parapluies. Alors que Deleuze se consacre à l'étude des forces à l'intérieur d'un tableau ou d'un triptyque de Bacon, l'idée de Focillon et Warburg sur les tensions des formes concerne non seulement les esquisses d'un tableau mais

⁹ Contrairement à la Digital Art History, la stratégie proposée par la Technical Art History s'attache à l'étude des changements progressifs entre une esquisse et une autre en reconstruisant les hésitations, la stratigraphie des repentirs et la réalisation progressive de formes incertaines. À ce sujet, voir Cardinali (2019).

également les relations de forces *entre les tableaux*. Dans un article intitulé « Aby Warburg et la science sans nom » (1999), Giorgio Agamben examine la conception de Warburg sur la transmission et la survivance des formes. Agamben décrit *la mémoire comme la capacité de réagir à un événement sur une période de temps variable* (l'intervalle mentionné ci-dessus à propos du travail de Focillon).

La trace laissée par l'événement constitue une énergie potentielle qui est stockée et qui peut être réactivée et déchargée. Dans les images aussi, cette énergie est conservée, et peut ensuite être sélectionnée selon les nécessités et les aléas des périodes successives. Il peut en résulter la reprise d'un motif, par exemple un motif du paganisme à la Renaissance, comme dans l'exemple de Ghirlandaio et de la *Naissance de Jean-Baptiste*. En d'autres termes, les images sont considérées par Warburg comme des dynamogrammes, qui sont des formes ayant enregistré une force dynamique et devant à leur tour la transmettre. L'esthétique des dynamogrammes que Warburg vise à établir peut être décrite à travers les trois phases d'un processus expliquant la survivance des motifs : la polarisation, la dépolarisation (ou déconnexion) et la repolarisation. C'est le contact avec l'ère nouvelle qui permet la polarisation, qui peut changer ou même inverser totalement le sens original de la composition.

Alors que la Digital Art History recherche actuellement les similitudes et les répétitions de motifs dans les collections d'images, ce que Warburg recherchait à son époque était la survivance des forces, et non pas la répétition des motifs stabilisés. Par force, il faut ici entendre des *types d'énergies condensées dans les formes du passé qui peuvent être autrement polarisées dans des images plus récentes*. L'affirmation suivante de Georges Didi-Huberman est très explicite sur la question qui tenait à cœur à Warburg : « L'iconographie peut s'organiser en motifs, voire en types – mais les formules de pathos, elles, définissent un champ que Warburg pensait comme rigoureusement *trans-iconographique* » (2001, p. 145, je souligne).

En effet, dans une image, l'émotion contamine tout, de manière transversale par rapport à la forme et au fond, voire par rapport aux motifs.

3.2. Formes et techniques

La notion de forme de Focillon se concentre sur des caractéristiques fondamentales de l'image qui ne sont pas du tout prises en compte par l'analyse computationnelle, notamment en ce qui concerne la technique de production des images analysées.

L'une des faiblesses détectables dans les pratiques actuelles d'analyse automatique des images est précisément le manque d'attention à la résonance des formes avec le support qui les organise, à savoir l'indifférence au médium. Car, comme l'affirme Focillon, c'est le médium qui infléchit les formes. À présent, les analyses computationnelles des similitudes visuelles des images sont effectuées principalement en combinant une première classification au

moyen de métadonnées standards (date de production, auteur, genre) et une classification ultérieure basée sur la Feature Extraction et le regroupement des images selon leurs caractéristiques plastiques. Comme déjà esquissé, le principal défaut de la classification par métadonnées standard est que les images sont classées en fonction d'informations *ajoutées* à l'image (Manovich, 2015), via des lexicalisations qui ne correspondent pas aux caractéristiques qui identifient chaque image comme unique et qui ne respectent donc pas les formes d'organisation qui lui sont propres. En outre, la procédure qui applique un premier filtre et opère une première classification via des métadonnées standard peut préfigurer certains résultats ou restreindre le corpus qui sera analysé ultérieurement par le biais de la Feature Extraction. En revanche, le principal défaut de la Feature Extraction utilisée seule est que les similitudes entre les formes qui sont identifiées par les algorithmes ne sont pertinentes ni d'un point de vue historique, ni en termes des techniques utilisées.

Dans les deux méthodes mentionnées (les métadonnées et la Feature Extraction), l'attention portée par la Computer Science aux caractéristiques du support de l'image, ou même à sa technique, est très faible, alors que la technique est ce qui compte le plus dans une image, parce que *la forme est formée dans la matière*. À cet égard, Focillon considère que la technique résulte de la dynamique établie entre le type d'outil, le mouvement de la main et les caractéristiques du support. En effet, chaque matière aura une certaine finalité, voire une « certaine vocation formelle », tout comme la forme aura une « certaine vocation matérielle ». Cette double vocation est intégrée dans les gestes de production.

Dans ses descriptions très riches de la peinture à l'huile, Focillon affirme que l'espace même dépend de la matière, ainsi que la lumière elle-même : la lumière dépend de la matière car la peinture à l'huile peut se glisser sur la matière du support par coulées ou se poser avec fermeté, pénétrer plus ou moins en lui communiquant une qualité sèche ou une qualité grasse (Focillon, 1934, p. 5 et sv). Ces gestualités ont une influence sur les effets lumineux de transparence, de luminosité, de brillance, etc. :

La même forme *conserve sa mesure* mais change de qualité selon la matière, l'outil et la main. [...] Une forme sans son support n'est pas forme, et le support est forme lui-même. Il est donc nécessaire de faire intervenir l'immense variété des techniques dans la généalogie de l'œuvre d'art et de montrer que le principe de toute technique n'est pas inertie mais action. (1934, p. 25-26, je souligne)

Focillon assimile ainsi la technique pas tant à la syntaxe qu'à une *physiologie* (*Ibid.*, p. 57) et à un *développement biologique* (*Ibid.*, p. 58).

Je tiens à souligner qu'en sémiotique, certains travaux ont récemment consacré une grande attention à la question des supports des œuvres d'art et

des inscriptions sur ces supports. Dans *Sémiotique de la photographie* (Basso Fossali et Dondero, 2011), nous avons étudié la relation entre quatre dispositifs expliquant le fonctionnement d'un médium :

- 1) un support matériel qui reçoit les inscriptions, par exemple la toile ;
- 2) une médiation formelle, c'est-à-dire l'organisation syntagmatique et la taille de la surface qui sont choisies comme pertinentes pour ce support matériel afin de l'inscrire ;
- 3) un apport tel que la composante chromatique par exemple, qui permet l'inscription de lignes et de traces sur un support matériel ;
- 4) le geste d'inscription, le rythme par lequel ces inscriptions peuvent être déposées et travaillées sur un support matériel.

Il est nécessaire de concevoir les formes représentées dans une image comme le résultat des forces impliquées dans chacune de ces composantes. Le support matériel, par exemple, est nécessairement caractérisé par des forces autonomes qui identifient sa spécificité, comme les qualités physiques du bois, du plastique, etc. Toutes les matières présentent des caractéristiques diverses de résistance, d'élasticité, de dureté, d'endurance, etc. qui peuvent être considérées comme des forces qui vont s'opposer à d'autres forces, de manière plus ou moins conflictuelle, par exemple les forces caractérisant le rythme gestuel lors de l'acte d'inscription de marques sur le support.

4. Le contraste entre les approches de l'histoire de l'art et celles de la *Computer Vision*

Le problème aujourd'hui, à la lumière des percées de la *Computer Vision*, peut être décrit comme suit : alors que Focillon et Warburg cherchent à étudier la métamorphose des formes, et donc aussi la relation écologique entre figure et fond, dans le contexte du Deep Learning, l'un des plus importants théoriciens des réseaux de neurones convolutifs (CNN), Yann Le Cun, affirme par contre, dans un article écrit en collaboration, donner la priorité aux éléments locaux de l'image :

Tout d'abord, dans les données matricielles telles que les images, les groupes de valeurs locales sont souvent fortement corrélés, formant des motifs locaux distinctifs qui sont facilement détectables. Deuxièmement, les statistiques locales des images et autres signaux sont invariables en fonction de l'emplacement. En d'autres termes, *si un motif apparaît dans une partie de l'image, il peut apparaître n'importe où, d'où l'idée d'unités situées à différents endroits*, partageant les mêmes poids et identifiant le même motif dans différentes parties du tableau. (Le Cun *et al.*, 2015, p. 439, je traduis)

Le fait que, selon Le Cun *et al.*, les motifs sont considérés comme indifférents à leur emplacement dans un ensemble donné implique que la complexité n'est pas prise en compte.

Certes, Le Cun et ses co-auteurs visent autre chose que l'analyse des œuvres d'art ; ils visent plutôt à reconnaître des objets représentés dans le contexte de compositions désordonnées et à produire automatiquement des légendes. Mais la reconnaissance des objets dans une image, artistique ou autre, ne peut constituer le seul objectif : dans une image, tout objet dépend d'un point de vue.

À ce propos, l'œuvre du philosophe René Thom est très utile car il aborde l'image non pas comme un ensemble de parties isolées mais comme une composition qui émerge de la tension entre des forces centrifuges et centripètes en tension. Dans son article sur l'art et l'esthétique, ayant pour titre « Local et global dans l'œuvre d'art » (1983), René Thom affirme que l'esthétique traite de la relation entre le local et le global. La composition d'une totalité dépend des relations entre les fragments perceptuels d'un tableau en relation à son cadre.

Thom aborde les tensions entre les forces au sein d'un tableau, et plus spécifiquement la radiation, la prolifération, etc. qui font du tableau une unité stable et compacte. Selon Thom, l'une des caractéristiques de la beauté dans la peinture est l'effet de contour. La beauté peut être mise en évidence par le cadre, qui permet d'identifier les centres de l'attention qui organisent la perception :

Dans l'observation d'un tableau (ou plus généralement d'une œuvre plastique) l'esprit commence à cerner le contour de l'œuvre ; puis, dans un effort d'analyse, on va s'efforcer de discerner, à l'intérieur des centres, des sujets porteurs d'une certaine prégnance. L'espace total de l'œuvre se trouve ainsi découpé en domaines partiels, qui sont *les zones de rayonnement d'un centre* (ou plus généralement d'une configuration locale de détails prise comme un individu). On peut penser que ce découpage provient d'une *sorte de prolifération du contour vers l'intérieur*, prolifération plus rapide là où aucun détail particulier ne retient l'attention [...]. C'est essentiellement le conflit de ces prégnances, régi par le logos d'une « catastrophe » (peut-être plus qu'élémentaire !) qui va assurer l'unité de l'œuvre d'art, son effet (1983, p. 5, je traduis)

Il est très important de souligner que l'unité de l'œuvre d'art est assurée par le conflit entre des forces contraires. Ce sont en effet les forces contraires qui entraînent la compacité d'un tableau ; là où il n'y a pas de conflit, il n'y a pas de cohésion, pas de composition.

Thom, comme Warburg, n'étudie pas seulement les forces internes à l'image, c'est-à-dire les forces qui se développent entre le contour et les centres de l'attention au sein d'une surface bidimensionnelle, mais aussi la propagation de la force de l'image vers l'extérieur, vers la corporité de l'observateur, par exemple.

En effet, selon Thom, la beauté est toujours localisée, et c'est le contour, ou le cadre, qui permet à la beauté d'émerger. Mais la beauté qui, pour exister, doit être encadrée, n'est pas nécessairement synonyme de limite et de clôture – bien au contraire car « un objet d'art est source d'une aura de beauté qui remplit tout son voisinage (selon des lois d'ailleurs fort subtiles car la propagation est loin d'être isotrope) » (*Ibid.*, p. 5). Il continue ainsi :

Une forme, par elle-même, suscite toujours une interprétation mécanique, un « champ de force ». Ce champ de force peut être d'origine subjective (selon la théorie de Harry Blum, *la reconnaissance d'une forme n'est autre que le choix d'une stratégie motrice optimale pour saisir – manuellement – cette forme*) ; elle peut être objective, décrivant les forces que l'objet peut émettre ou subir. (*Ibid.*, p. 7, je souligne)

Ce qui est pertinent ici, c'est la relation entre les forces à l'intérieur d'une image et les stratégies motrices pour appréhender ces forces (stratégies que Thom appelle subjectives), ainsi que ce qu'il considère comme étant d'origine objective, c'est-à-dire le fait que l'objet pictural reçoit des forces émanant de la matière physique, comme la lumière, qu'il fixe et stabilise sur un substrat matériel, et qu'il redirige ensuite vers l'extérieur, y compris vers d'autres images et vers les observateurs.¹⁰

5. Conclusion

Même si un effort a déjà été fait dans le traitement des images pour passer de descriptions basées sur des métadonnées standards (qui dépendent de la segmentation verbale du langage visuel) à l'extraction de caractéristiques plastiques, il faudrait aller plus loin et orienter le Deep Learning vers une approche méréologique des tensions entre les forces centrifuges et centripètes. Si, de son côté, Warburg était surtout intéressé par la recherche de liens entre les images du point de vue du *Pathosformel* et par la manière dont les émotions s'expriment dans l'image à travers la gestualité corporelle, le travail de Focillon nous incite à faire progresser les ambitions de l'analyse sémiotique concernant la substance du plan de l'expression, ainsi qu'à pallier à certaines faiblesses de l'analyse sémiotique classique (analyse des corpus, de la technique de production des images, de la diachronie, l'intertextualité). En effet, Focillon non seulement introduit le vecteur de transformation dans le temps des styles et des techniques de production d'images, mais il rend également justice à l'image en tant que totalité indivisible, à la stratification des esquisses, et à la spécificité des médias.

¹⁰ Sur la théorie de Thom et les émotions liées à l'œuvre d'art, voir Dondero (2020b).

Références bibliographiques

- Agamben Giorgio, 1999, « Aby Warburg and the Nameless Science », *Potentialities*, Agamben Giorgio (dir.), Stanford, Stanford University Press, p. 89-103.
- Basso Fossali Pierluigi, Dondero Maria Giulia, 2011, *Sémiotique de la photographie*, Limoges, Presses Universitaires de Limoges.
- Bishop Claire, 2018, « Against Digital Art History », *International Journal for Digital Art History*, n° 3, disponible en ligne.
- Cardinali Marco, 2019, « Digital Tools and Technical Views: The Intersection of Digital Art History and Technical Art History in a Digital Archive on the Painting Technique of Caravaggio and His Followers », *Visual Resources*, n° 35, p. 52-73.
- Colas-Blaise Marion, 2019, « Comment penser la narrativité dans l'image fixe ? La "composition cinétique" chez Paul Klee », *Pratiques*, n° 181-182, disponible en ligne.
- Deleuze Gilles, 1981, *Francis Bacon. Logique de la sensation*, Paris, Différence.
- Didi-Huberman Georges, 2001, « Aby Warburg et l'archive des intensités », *Études photographiques*, n° 10, p. 144-160, disponible en ligne.
- Dondero Maria Giulia, 2017, « The Semiotics of Design in Media Visualization: Mereology and Observation Strategies », *Information Design Journal*, vol. 23, n° 2, p. 208-218.
- Dondero Maria Giulia, 2019, « Visual Semiotics and Automatic Analysis of Images from the Cultural Analytics Lab: How Can Quantitative and Qualitative Analysis be Combined? », *Semiotica*, n° 230, p. 121-142, disponible en ligne.
- Dondero Maria Giulia, 2020a, *Les Langages de l'image : de la peinture aux Big Visual Data*, Paris, Hermann.
- Dondero Maria Giulia, 2020b, « Les forces dans l'image et les gestualités émotionnelles », *Émotion, empathie, affectivité : les sujets et leur subjectivité à travers les pratiques langagières et éducatives*, Acosta Córdoba Luisa Fernanda, Piccoli Vanessa (dirs.), Les Ulis, EDP Sciences, p. 1-14, disponible en ligne.
- Fabbri Paolo, 1998, *La svolta semiotica*, Rome-Bari, Laterza [traduction française : *Le Tournant sémiotique*, Paris, Lavoisier, 2008].
- Floch Jean-Marie, 1985, *Petites mythologies de l'œil et de l'esprit : pour une sémiotique plastique*, Paris/Amsterdam, Hadès-Benjamins.
- Focillon Henri, [2010] 1934, *Vie des formes*, Paris, Presses Universitaires de France.
- Fontanille Jacques, 1995, *Sémiotique du visible : des mondes de lumière*, Paris, Presses Universitaires de France.

- Fontanille Jacques, 1998, « Décoratif, iconicité et écriture : geste, rythme et figurativité. À propos de la poterie berbère », *Visio*, vol. 3, n° 2, disponible en ligne.
- Fontanille Jacques, Zilberberg Claude, 1998, *Tension et signification*, Sprimont, Mardaga.
- Greimas Algirdas Julien, 1984, « Sémiotique figurative et sémiotique plastique », *Actes Sémiotiques*, n° 60, disponible en ligne.
- Hagelstein Maud, 2009, « Mnemosyne et le Denkraum renaissant : pratique du document visuel chez Aby Warburg », *MethIS*, n° 2, p. 87-111.
- Hall Gary, 2013, « Toward a Postdigital Humanities: Cultural Analytics and the Computational Turn to Data-Driven Scholarship », *American Literature*, vol. 85, n° 4, p. 781-810.
- Hristova Stefka, 2016, « Images as Data: Cultural Analytics and Aby Warburg's Mnemosyne », *International Journal for Digital Art History*, n° 2, disponible en ligne.
- Journal of Knowledge, Culture and Communication*, n° 36, disponible en ligne.
- Le Cun Yann *et al.*, 2015, « Deep learning », *Nature*, n° 521, p. 436-444.
- Manovich Lev, 2011, « Style Space: How to Compare Image Sets and Follow their Evolution », disponible en ligne.
- Manovich Lev, 2015, « Data Science and Digital Art History », *International Journal for Digital Art History*, vol. 1, n° 1, p. 3-35.
- Manovich Lev, 2017, « The Science of Culture? Social Computing, Digital Humanities and Cultural Analytics », *The Datafied Society: Studying Culture through Data*, Schäfer Mirko, van Es Karin (dirs.), Amsterdam, AUP, p. 55-68.
- Manovich Lev, 2020, « Computer Vision, Human Senses, and Language of Art », *AI & Society*,
- Manovich Lev, Douglass Jeremy, 2009, « Timeline. 4535 Time Magazine Covers, 1923- 2009 », *Cultural Analytics*, disponible en ligne.
- Massaert Lucien, 2019, « Immanence du savoir et configuration du tableau », *Signata*, n° 10, disponible en ligne.
- Seguin Benoit, 2018, « Making Large Art Historical Photo Archives Searchable », *Dissertation, EPFL Scientific Publications*, disponible en ligne.
- Thom René, 1983, « Local et global dans l'œuvre d'art », *Le Débat*, vol. 24, n° 2, p. 73-89.
- Thom René, 2006, « L'arte : luogo di conflitto tra forme e forze ? », *Morfologia del semiotico*, Fabbri Paolo (dir.), Rome, Meltemi, p. 133-142.
- Warburg Aby, 2012, *L'Atlas Mnemosyne*, Paris, Atelier de l'écarquillé.
- Zilberberg Claude, 2012, *La Structure tensive*, suivi de *Note sur la structure des paradigmes* et de *Sur la dualité de la poétique*, Liège, Presses Universitaires de Liège.